

Міністерство освіти і науки, молоді та спорту України  
Харківський національний університет  
імені В. Н. Каразіна

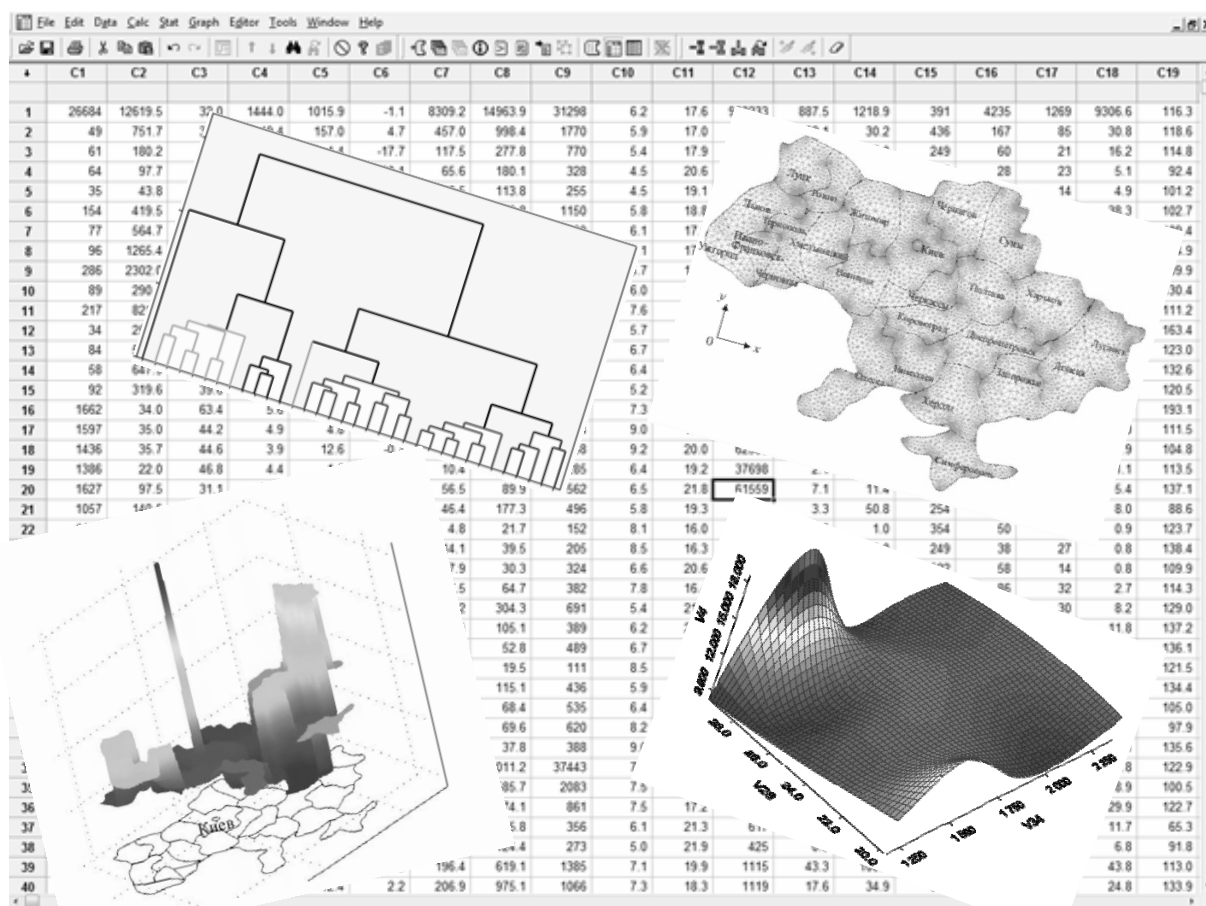


Геолого-географічний факультет  
Кафедра соціально-економічної географії  
і регіонознавства



# СТАТИСТИЧНІ МЕТОДИ І ОБРОБКА ГЕОІНФОРМАЦІЇ

Навчально-методичний комплекс для самостійної роботи  
студентів, які навчаються за напрямом підготовки «Географія»  
зі спеціальності «Економічна та соціальна географія»



Харків – 2012

УДК 911.3:30

ББК 65.04

Н50

*Схвалено методичною комісією геолого-географічного факультету  
Харківського національного університету імені В. Н. Каразіна  
(протокол № 7 від 17.05.2012 р.)*

*Рекомендовано до друку Вченою радою геолого-географічного факультету  
Харківського національного університету імені В. Н. Каразіна  
(протокол № 8 від 20.05.2012 року)*

**Н50 Нємець К.А., Сегіда К.Ю. Статистичні методи і обробка геоінформації:** навчально-методичний комплекс для самостійної роботи студентів, які навчаються за напрямом підготовки «Географія» зі спеціальності «Економічна та соціальна географія». – Харків, ХНУ ім. В.Н. Каразіна, 2012. – 52 с.

Навчально-методичний комплекс розроблений відповідно до програми курсу «Статистичні методи і обробка геоінформації» як одного з важливих у загальному циклі суспільно-географічної та загальноекономічної підготовки фахівців.

Мета: надати методичну допомогу студентам та розвивати навички самостійної роботи при вивченні курсу «Статистичні методи і обробка геоінформації».

Навчально-методичний комплекс розрахований на студентів денного і заочного відділень геолого-географічного факультету, які навчаються за спеціальністю «Економічна і соціальна географія». Вони містять загальні відомості про курс, тематичний план (структуру) курсу, навчальну програму з вимогами до формування компетентності студентів, рекомендовану літературу, перелік тем і планів практичних та семінарських занять, основні положення та методичні вказівки до їх виконання, теми індивідуальних дослідних завдань, питання до модульного контролю, систему оцінювання навчальних досягнень студентів, а також короткий глосарій основних понять курсу.

УДК 911.3:30

ББК 65.04

Н50

© Харківський національний університет  
імені В. Н. Каразіна, 2012

© Нємець К.А., Сегіда К.Ю., 2012

## ЗМІСТ

ЗАГАЛЬНІ ПОЛОЖЕННЯ КУРСУ .....	4
СТРУКТУРА І РОБОЧИЙ ПЛАН КУРСУ .....	6
ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ.....	7
РЕКОМЕНДОВАНА ЛІТЕРАТУРА.....	9
ТЕМИ СЕМІНАРСЬКО-ПРАКТИЧНИХ ЗАНЯТЬ.....	11
Практично-семінарське заняття 1: .....	12
Практично-семінарське заняття 2: .....	12
Практично-семінарське заняття 3: .....	17
Практично-семінарське заняття 4: .....	20
ПИТАННЯ ДО ПІДСУМКОВОГО МОДУЛЬНОГО КОНТРОЛЮ.....	33
Приклад підсумкового модульного контролю .....	35
ОЦІНЮВАННЯ РЕЗУЛЬТАТІВ НАВЧАЛЬНИХ ДОСЯГНЕНЬ СТУДЕНТІВ	38
КОРОТКИЙ ГЛОСАРІЙ.....	39

## ЗАГАЛЬНІ ПОЛОЖЕННЯ КУРСУ

Одночасно із стрімким розвитком науки, інтеграцією окремих її галузей, активно розвивається методологія суспільної географії. Новий соціальний запит формує необхідність комплексних досліджень, зумовило впровадження в суспільну географію математико-статистичних методів досліджень. Для підготовки фахівців в галузі соціальної та економічної географії у ВНЗ викладається курс «Статистичні методи і обробка геоінформації», який є спеціальним у підготовці бакалаврів географії та магістрів (спеціалістів) соціально-економічної географії. Він поєднує загальнонаукові підходи (математико-статистичний, математичний) з методологічними основами суспільної географії, що дає можливість створити у студента компетентність стосовно обробки статистичних даних, побудови статистичних моделей і коректної обробки геоінформаційних даних. Курс «Статистичні методи і обробка геоінформації» викладається паралельно з курсом «Теорія і методологія географічної науки», що дозволяє студентами зрозуміти роль і місце статистичних методів у географічній методології. В той же час, курс передує багатьом іншим спецкурсам для спеціальності «Економічна та соціальна географія», що створює сприятливі можливості для застосування статистичних методів у виконанні самостійних завдань з цих курсів.

**Мета курсу** – сформулювати у студентів компетентності стосовно застосування статистичних методів, математичного моделювання та комп'ютерних технологій в суспільно-географічних дослідженнях.

**Предметом курсу** є вивчення теоретичних основ теорії ймовірностей та математичної статистики як науки про дослідження випадкових процесів і обробку випадкових величин, а також про конкретні методи статистичного моделювання суспільно-географічного процесу.

### **Завдання курсу:**

- сформулювати у студентів сучасну методологію використання математичних методів, моделей та комп'ютерних технологій при дослідженні соціально-географічних систем;

- дати студентам знання та поняття стосовно основних методів і підходів в математичній обробці суспільно-географічної інформації;

- сформулювати у студентів поняття про математичні методи та моделювання при вирішенні суспільно-географічних задач;

- сформулювати у студентів компетентність стосовно використання комп'ютерних технологій та математичних моделей в суспільно-географічних дослідженнях.

У результаті вивчення даного курсу:

**студенти мають знати поняття:** спостереження, опис, вимірювання, методологія, гносеологія, теоретичне знання, емпіричне знання, парадигми науки, практика, прогноз, пізнання, модель, математична модель, геосистемна, синергетична парадигма, інформаційно-синергетична парадигма, ідеалізація, формалізація, абстрагування, синтез понять, ймовірність подій, закон великих чисел, статистична стійкість, системний аналіз, вибірковий метод, причинно-наслідкові зв'язки, локальні процеси, інтегральні процеси, погано організовані системи, добре організовані системи, детерміновані процеси, випадкові процеси, детерміновані моделі, ймовірнісні моделі, частота, частість, розподіл випадкової величини, гістограми, полігони частот, центр розподілу, математичне очікування, дисперсія, стандарт, асиметрія, ексцес, нормальний розподіл, логарифмічно нормальний розподіл, лінія регресії, коефіцієнти регресії коваріація, коефіцієнт кореляції, множинний і частковий коефіцієнт кореляції, графи, дендрограми, розпізнання образів, власна область, гіперповерхня, узагальнена відстань, специфічність та загальність змінних, згортка інформації, поле просторової змінної, закономірна та випадкова мінливість, тренд-аналіз, фон, аномалія, генералізація фона, алгебраїчні та тригонометричні поліноми.

**студенти мають вміти:** планувати суспільно-географічні дослідження в залежності від системи цілей, можливостей виконавської системи, конкретних умов, розкладати складні процеси на прості, описати причинно-наслідкові зв'язки, визначити ймовірнісний характер процесів, розрізняти локальні та інтегральні процеси, оцінювати ймовірність події, описати закон розподілу випадкової величини, застосовувати методи статистичного аналізу і графічно відображати результати статистичного дослідження за допомогою комп'ютера, використовувати стандартні пакети прикладних програм статистичного аналізу для вирішення конкретних прикладних задач суспільної географії, використовувати просторові змінні для опису та аналізу географічних полів.

## СТРУКТУРА І РОБОЧИЙ ПЛАН КУРСУ

Назви модулів і тем	Кількість годин											
	Денна форма						Заочна форма					
	Усього	у тому числі					Усього	у тому числі				
		л	п	лаб	інд	ср		л	п	лаб	інд	ср
1	2	3	4	5	6	7	8	9	10	11	12	13
<b>Модуль 1. Особливості використання математичних методів (моделювання) в суспільній географії</b>												
<b>Тема 1.</b> Місце математичної статистики в географічній методології	26	4	2			20	18	2				16
<b>Тема 2.</b> Особливості природний і суспільних процесів, як об'єктів статистичного моделювання	26	4	2			20	18	2				16
<i>Разом за модулем 1</i>	52	8	4			40	40	4				36
<b>Модуль 2. Статистичне моделювання</b>												
<b>Тема 1.</b> Одновимірні статистичні моделі та їхні особливості	19	4	2			13	14	2				12
<b>Тема 2.</b> Двовимірні статистичні моделі і особливості їх застосування	22	4	2			16	18	2				16
<b>Тема 3.</b> Багатовимірні статистичні моделі	22	4	2			16	16	2	2			12
<b>Тема 4.</b> Моделювання просторових змінних	29	6	3			20	20	2	2			16
<i>Разом за модулем 2</i>	92	18	9			65	68	8	4			56
Усього годин	144	26	13			66	108	12	4			92

# **ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**

## **Модуль 1. Особливості використання математичних методів (моделювання) в суспільній географії**

### **Тема 1. Місце математичної статистики в географічній методології**

*Лекція 1. Основні поняття та історія використання статистичних методів у суспільній географії. Об'єкт та мета курсу. Визначення математичних методів та моделювання. Поняття «модель». Класифікації моделей: натурні, аналогові, математичні. Історія розвитку і використання моделей в природознавстві і суспільних науках.*

*Лекція 2. Моделювання, як метод пізнання, в суспільній географії. Три рівні математизації суспільної географії: рівень параметризації географічних явищ та об'єктів, рівень емпіричних моделей, рівень теоретичних моделей. Місце методів ідеалізації, формалізації, математичних методів, методів моделювання в методології суспільно-географічного дослідження. Протиріччя системного і синергетичного підходів. Перспективи розвитку моделювання в суспільній географії.*

### **Тема 2. Особливості природний і суспільних процесів, як об'єктів статистичного моделювання**

*Лекція 3. Специфіка розвитку природних і соціальних об'єктів. Специфіка виникнення, функціонування і розвитку природних та соціальних об'єктів. Багатофакторність суспільно-географічних процесів. Складність інтерпретації причинно-наслідкових зв'язків в соціогеосистемах.*

*Лекція 4. Випадковий характер організації географічного середовища. Організація географічного середовища. Локальні та інтегральні процеси. Поняття про ймовірність. Принципово випадковий характер суспільно-географічних процесів. Особливості суспільно-географічної інформації.*

## **Модуль 2. Статистичне моделювання**

### **Тема 1. Одновимірні статистичні моделі та їхні особливості**

*Лекція 5. Сутність одновимірних статистичних моделей. Одновимірні статистичні моделі. Поняття про випадкову величину, дискретні і неперервні випадкові величини. Числові характеристики випадкових величин: частота, частість, функція розподілу. Оцінки центру розподілу: математичне очікування,*

медіана, мода. Оцінки розкиду випадкової величини: центральні моменти другого, третього і четвертого порядків. Основні закони одновимірного розподілу.

*Лекція 6. Оцінка параметрів суспільно-географічних об'єктів.* Вибірковий метод, вимоги до вибірових сукупностей. Поняття про надійну ймовірність. Точкова та інтервальна оцінка статистик. Вимоги до оцінок параметрів суспільно-географічних об'єктів. Поняття про перевірку статистичних гіпотез. Умови застосування одновимірних статистичних моделей.

## **Тема 2. Двовимірні статистичні моделі і особливості їх застосування**

*Лекція 7. Двовимірні статистичні моделі.* Двовимірні статистичні моделі. Поняття про двовимірну випадкову величину. Умови формування двовимірних випадкових величин у суспільно-географічних дослідженнях. Особливості формування вибірових сукупностей. Особливості інтерпретації зв'язків компонентів двовимірних випадкових величин. Поняття про форму залежності і силу зв'язку двох випадкових величин. Функціональні і статистичні залежності.

*Лекція 8. Регресійно-кореляційний аналіз та його застосування.* Двовимірний регресійний аналіз, поняття про умовний розподіл, лінійні, нелінійні рівняння регресії, обчислення регресійних коефіцієнтів. Побудова довірчого інтервалу для рівнянь регресії. Кореляційний аналіз, поняття про коефіцієнт кореляції, кореляційне відношення. Умови коректності кореляційного аналізу. Нелінійні перетворення випадкових величин для приведення до нормального закону розподілу. Особливості використання регресійно-кореляційного аналізу у суспільно-географічних дослідженнях.

## **Тема 3. Багатовимірні статистичні моделі**

*Лекція 9. Багатовимірний статистичний аналіз.* Багатовимірні статистичні моделі. Поняття про багатовимірну випадкову величину. Поняття про матрицю вихідних даних. Багатовимірний кореляційний аналіз, парний, частковий, множинний коефіцієнт кореляції. Особливості використання багатовимірного кореляційного аналізу. Багатовимірний регресійний аналіз, визначення значущості часткових коефіцієнтів регресії. Особливості застосування багатовимірного регресійного аналізу.

*Лекція 10. Методи багатовимірного статистичного аналізу, особливості їх застосування.* Поняття про багатовимірний ознаковий простір, визначення відстані у багатовимірному просторі. Використання теорії графів для класифікації суспільно – географічних об'єктів. Кластер-аналіз, визначення дистанційних коефіцієнтів, принципи утворення кластерів. Задачі розпізнання образів, поняття про власну область об'єкту, вирішальну функцію, вирішальне правило.



Алгоритм розпізнавання образів. Дискримінантний аналіз у задачах класифікації. Факторний аналіз, його різновиди, інтерпретація результатів.

#### **Тема 4. Моделювання просторових змінних**

*Лекція 11. Просторові змінні як об'єкт геоінформатики.* Поняття про просторову змінну, ознаки і властивості просторових змінних. Відмінності моделей просторових змінних від статистичних моделей. Суспільно-географічні об'єкти як генератори полів просторових змінних. Мінливість та анізотропність полів просторових змінних. Закономірна та випадкова складові мінливості, що спостерігається.

*Лекція 12. Методи розділення фону та аномалій, методи розрахунку локальних показників.* Поняття про фон та аномалію. Поняття про тренд. Тренд-аналіз як метод розділення фону та аномалій. Задачі тренд-аналізу в суспільній географії. Загальна характеристика методів згладжування випадкових полів. Метод ковзного статистичного вікна. Принципи розрахунку локального середнього. Розмір ковзного статистичного вікна як параметр генералізації карти апроксимації випадкового поля. Аналіз залишків та виділення аномалій. Інтерпретація результатів тренд-аналізу.

*Лекція 13. Апроксимація поверхні тренда єдиною функцією координат простору.* Загальний огляд методів апроксимації поверхні тренда єдиною функцією координат простору. Принципи оцінки апроксимації поверхні тренду. Моделі алгебраїчних поліномів, вибір порядку поліному. Моделі тригонометричних поліномів. Використання диференційних рівнянь для апроксимації поверхні тренду.

#### **РЕКОМЕНДОВАНА ЛІТЕРАТУРА**

1. Арманд А.Д. Информационные модели природных комплексов. – М.: Наука, 1975.
2. Арманд А.Д. Самоорганизация и саморегулирование географических систем. – М.: Наука, 1988.
3. Бендат Дж., Пирсон А. Прикладной анализ случайных данных. М., Мир, 1989.
4. Браверман Э.М., Мучник И.Б. Структурные методы обработки эмпирических данных. М., Наука, 1983.
5. Бусленко Н.П. Моделирование сложных систем. М., Наука, 1976.
6. Вайдлих В. Социодинамика. Системный подход к математическому моделированию в социальных науках. – М.: Эдиториал УРСС, 2004.

7. Веников В.А. Теория подобия и моделирования (применительно к задачам электроэнергетики). – М.: Высшая школа, 1976.
8. Гродзинський М.Д. Стійкість геосистем до антропогенних навантажень. – Київ: Ліцей, 1995.
9. Гроп Д. Методы идентификации систем. Пер. с англ. В.А. Васильева, В.И. Лопатина. – М.: Мир, 1979.
10. Ивахненко А.Г., Мюллер И.А. Самоорганизация прогнозирующих моделей. Киев, Техника, 1985.
11. Исаченко А.Г. Оптимизация природной среды (географический аспект). – М., 1980.
12. Каждан А.Б., Гуськов О.И. Математические методы в геологии. М., Недра, 1990.
13. Капица С.П. Общая теория роста человечества. Сколько людей жило, живет и будет жить на Земле. – М.: Наука, 1999.
14. Математическое моделирование: проблемы и результаты. – М.: Наука, 2003.
15. Моделирование динамики геосистем регионального уровня. Хомяков П.М. и др. – М.: Изд – во МГУ, 2000.
16. Моисеев Н.Н. Математические задачи системного анализа. М., Наука, 1981.
17. Немець Л.М., Немець К.А. До методики соціально – географічного асоціативного аналізу // Економічна та соціальна географія: міжвід. наук. зб. – К., 2003, – вип. 54. – с. 13 – 18.
18. Немець Л.М., Олійник Я.Б., Немець К.А. Просторова організація соціально – географічних процесів в Україні. – Київ – Харків: РВВ ХНУ, 2003.
19. Осауленко О.Г. Моделювання сталого розвитку соціально-економічних систем: Монографія. – К., 2001.
20. Пістун М.Д. Основи теорії суспільної географії. – К.: Вища школа, 1996.
21. Поздняков А.В., Черванев И.Г. Самоорганизация в развитии форм рельефа. – М.: Наука, 1990, 204 с.
22. Пригожин И., Стенгерс И. Порядок из хаоса. Новый диалог человека с природой. – М.: Эдиториал УРСС, 2001.
23. Пэнтл Р. Методы системного анализа окружающей среды. М., Мир, 1979.
24. Самарский А.А., Гулин А.В. Численные методы. М., Наука, 1989.
25. Сочава В.Б. Введение в учение о геосистемах. – Новосибирск: Наука, 1978.
26. Топчієв О.Г. Суспільно – географічні дослідження: методологія, методи, методики. Навч. посіб. – Одеса: Астропринт, 2005.
27. Хакен Г. Информация и самоорганизация. Макроскопический подход к сложным системам. – М.: Мир, 1991.
28. Харвей Дж. Модели развития пространственных систем в географии человека / / Модели в географии. – М., 1971.

29. Шаблій О.І. Математичні методи в соціально-економічній географії: Навч. видання. – Львів: Світ, 1994.
30. Шаблій О.І. Суспільна географія: теорія, історія, українознавчі студії. – Львів: Львівський національний університет імені Івана Франка, 2001.
31. Шаблій О.І. Основи загальної суспільної географії. – Львів, 2003.
32. Шредингер Э. Пространственно-временная структура Вселенной. – Новокузнецк, ИО НФМИ, 2000.
33. Эткинс П. Порядок и беспорядок в природе. Пер. с англ. Ю.Г. Рудого. – М.: Мир, 1987.

## ТЕМИ СЕМІНАРСЬКО-ПРАКТИЧНИХ ЗАНЯТЬ

№ з/п	Назва теми	Кількість годин
1	Усвідомлення суспільно–географічної і математичної суті поставленої задачі, підготовка вихідної матриці даних.	2
2	Побудова одновимірних статистичних моделей векторів змінних.	3
3	Побудова двовимірних статистичних моделей на матриці вихідних даних.	4
4	Кластер-аналіз та факторний аналіз вихідних даних. Інтерпретація отриманих результатів.	4
	<b>Разом</b>	<b>13</b>

Семінарсько-практичні заняття з курсу побудовані таким чином, щоб сприяти послідовному засвоєнню теоретичного матеріалу з курсу, застосуванню на практиці математико-статистичних методів. Досягнення поставленої мети планується шляхом послідовного виконання таких завдань:

- Визначення об’єкту та предмету дослідження, формування бази даних;
- Статистична перевірка обраних показників (одновимірний аналіз);
- Визначення взаємозв’язків та взаємозалежності між показниками (двовимірний аналіз);
- Визначення територіальних та структурних особливостей суспільно-географічного процесу (кластер- та фактор-аналізи);
- Інтерпретація результатів, отриманих шляхом математико-статистичного дослідження.

## **Практично-семінарське заняття 1:**

### **Усвідомлення суспільно-географічної і математичної суті поставленої задачі, підготовка вихідної матриці даних**

1. Визначення об'єкту та предмету дослідження
2. Поняття «матриці даних»
3. Вимоги до складання матриці даних
4. Відбір статистичних показників для аналізу
5. Формування вихідної матриці даних

Першим завданням практичної роботи є статистичне спостереження – масове чи вибіркове, планомірне, науково-організоване спостереження за масовими явищами, що полягає в реєстрації відібраних ознак у кожної одиниці вибіркової сукупності.

Перший етап – це ознайомлення з теоретичним матеріалом курсу, наступний – проведення вибіркового збору соціально-економічних показників за статистичними щорічниками (України, Харківської або будь-якої іншої області, виходячи із обраного предмету дослідження). Статистичний показник – це узагальнююча кількісна характеристика властивостей сукупності загалом чи її частин зокрема щодо конкретних умов місця і часу. Статистичні показники поділяються на абсолютні, відносні та середні величини. Якісний зміст показника визначається суттю явища і знаходить своє відображення у назві. Кількісну сторону представляють число та розмірність (вимірник). Дане спостереження є не суцільним, оскільки охоплює лише частину суспільно-економічних одиниць, які характеризують обрані райони.

Наступним кроком є побудова інформаційної бази даних за статистичною сукупністю – це впорядкована множина існуючих елементів (однорідних у певному відношенні), поєднаних спільними, притаманними кожному елементу множини, властивостями, умовами та причинами існування і розвитку. Формування якісної інформаційної бази – це фундамент статистичного дослідження, оскільки використання лише об'єктивної та достатньо повної інформації на подальших етапах дослідження надає можливість отримати правильні, обґрунтовані висновки про характер і закономірності досліджуваного процесу.

## **Практично-семінарське заняття 2:**

### **Побудова одновимірних статистичних моделей векторів змінних**

Другим завданням практичної роботи є виявлення залежності динаміки певних показників від часу, включають методи згладжування та аналітичного

вимірювання. Згладжування базується на процедурі визначення усередненої траєкторії розвитку показників у минулому та її продовженні на майбутнє. Аналітичне вимірювання передбачає підбір математичної функції, яка найкращим чином відображає тенденцію динаміки показника у часі, та розрахунок на його основі прогнозних значень даного показника. Виконується за допомогою одно-вимірний аналізу, який здійснюється у програмі MINITAB 14. Одновимірний аналіз відноситься до базових, виконується шляхом «Stat» → «Basic Statistics».

Одновимірний аналіз – містить одну випадкову величину (аналіз однієї випадкової величини). В основі даного аналізу є дослідження властивостей та характеристик рядів випадкових величин; встановлення приналежності цих рядів до певного теоретичного розподілу; визначення схожості або відмінності даного ряду порівняно з іншими рядами.

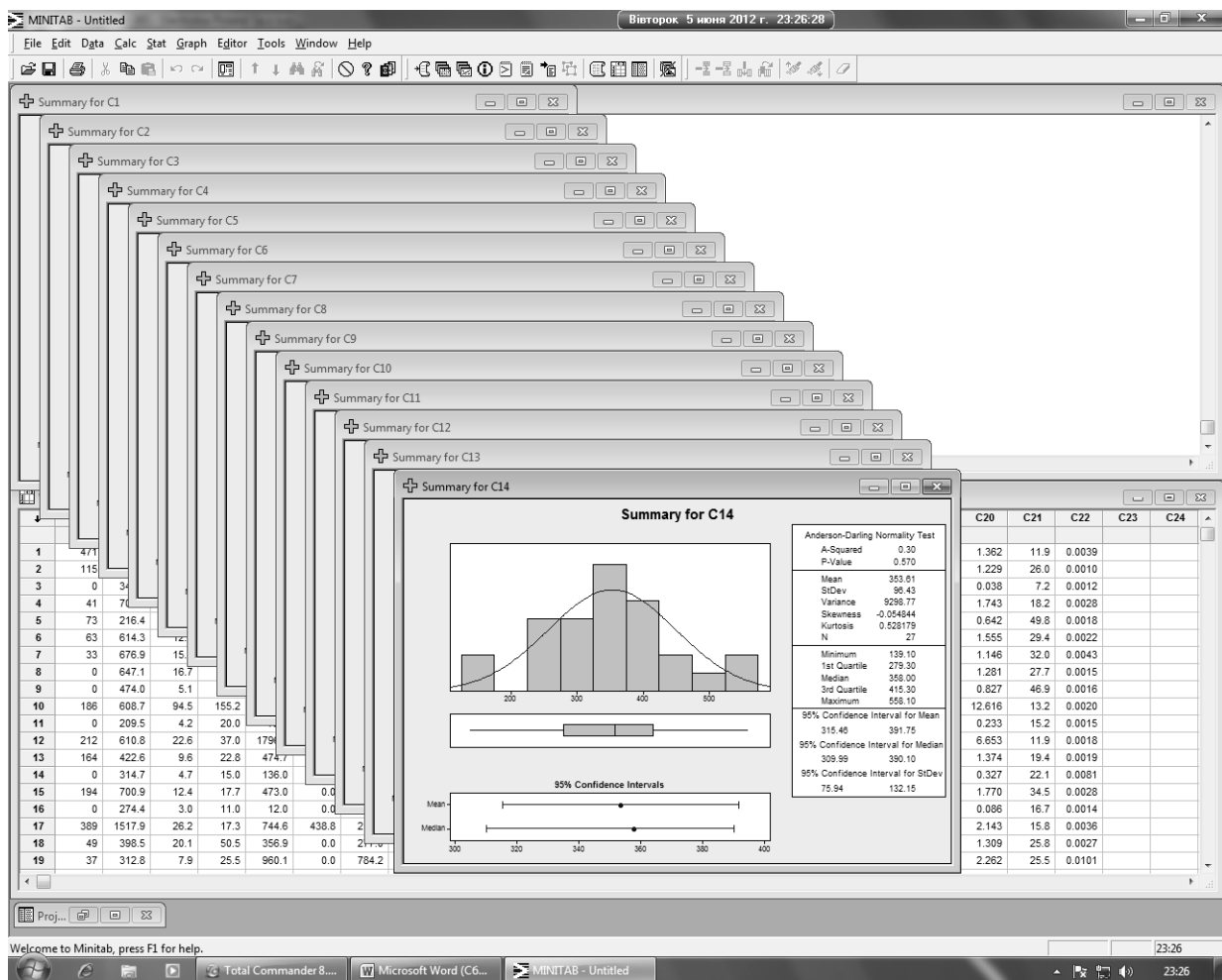


Рис. 1. Нормальний розподіл статистичного показника

Основними початковими складовими одновимірний аналізу є поняття: простий статистичний ряд (сукупність), впорядкований статистичний ряд (ранжируваний), варіаційний ряд, генеральна сукупність і вибірка, об'єм вибірки, частота і ймовірність, закон розподілу, критерії однорідності та інші.

Сукупність випадкових величин, отриманих за однакових умов проведення спостереження, дослідів, експериментів називається простою статистичною сукупністю або статистичним рядом. Статистичний ряд – це первинна форма запису статистичного матеріалу у вигляді таблиці. Для наочного представлення вихідного матеріалу за даними статистичного ряду будується графік змін значень даної величини в часі або в просторі або в хронологічній послідовності.

Генеральна сукупність – це нескінченне або кінцеве число елементів або компонентів, що складаються з якісно однорідних показників. Будь – яку частину генеральної сукупності, відібрану за певними правилами, яка характеризує генеральну сукупність, називають статистичною вибіркою. Вибіркову сукупність (вибірку) створюють звичайно для полегшення обробки інформації.

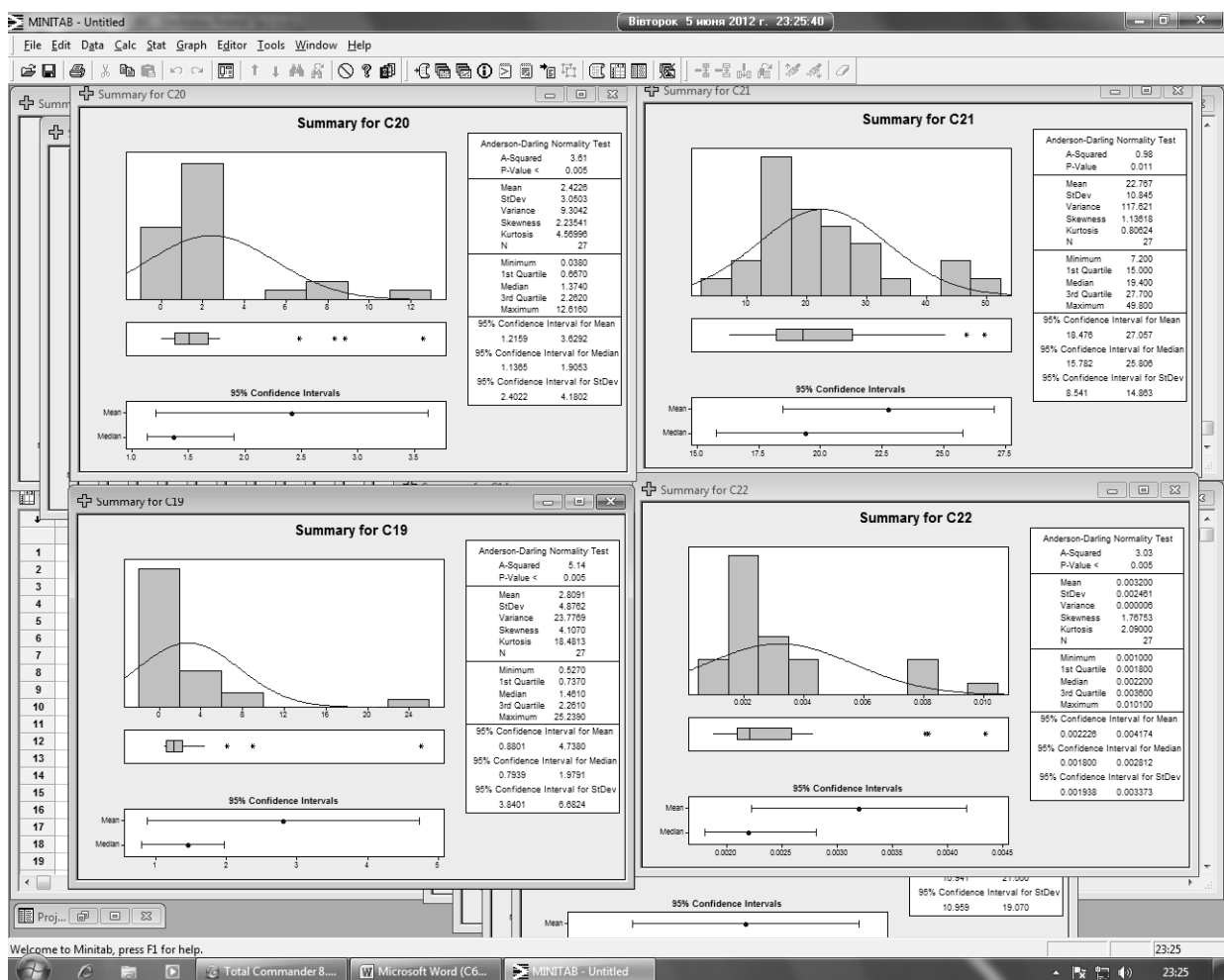


Рис. 2. Приклад одновимірного аналізу статистичних показників

Одним з найскладніших та найважливіших питань одновимірного статистичного аналізу є визначення кількості спостережень в дослідженнях для отримання надійного уявлення про характер мінливості ознаки, що вивчається, в генеральній сукупності. Зазвичай оптимальний об'єм вибірки пропорційний ступеню мінливості ознаки. Якщо ознака сильно змінюється, то кількість вимірю-

вань слід збільшити. Величину вибіркової сукупності при виконанні географічних досліджень можна визначити двома способами: за таблицею достатньо великих чисел; розрахунковим способом. У обох випадках кількість спостережень (чисельність чи об'єм вибірки) визначається, виходячи з довірчої ймовірності.

Систематизація і впорядкування даних, що представляють статистичну сукупність. Приведення їх в певну систему і характеристика цієї системи. Всяка статистична сукупність характеризується об'ємом ряду та абсолютною частотою повторення однакової ознаки. Частота – це число, яке показує скільки разів зустрічається дана ознака в сукупності, що вивчається. Закон зміни частоти – це і є закон розподілу. Існує три способи представлення закону розподілу: табличний, графічний і аналітичний. Результати обробки вихідних даних, як правило, спочатку завжди оформляються у вигляді статистичних таблиць. Така форма дозволяє надати матеріалу зручність, компактність та раціональність.

Спосіб, який дозволяє в наочній формі отримати уявлення про закономірності розподілу називається графічним зображенням варіаційного ряду. Існує декілька способів графічного зображення рядів розподілу. При цьому звичайно використовують дані таблиці емпіричного розподілу. При графічному зображенні рядів розподілу на горизонтальній осі відкладаються значення інтервалів або спостережені значення випадкової величини, а на вертикальній осі – частоти. Для наочного уявлення представлення варіаційного ряду найчастіше використовують графічні зображення у вигляді гістограми, полігону, кумуляти, кривої концентрації Лоренса та інші. Гістограма наочно показує розподіл величин, які вивчаються, через що подібний спосіб уже часто використовується для ілюстрації особливостей статистичного розподілу. Варіаційний ряд при цьому зображується у вигляді стовпчиків, границі між якими проходять по ординатах, відповідних границям між класами. При цьому основа стовпчиків по ширині дорівнює величині інтервалу ознаки, а висота – пропорційна частоті окремих класів. Розглянемо основні статистичні характеристики варіаційних рядів. Одним з основних параметрів статистичного ряду є середнє значення ознаки або центр, відносно якого розподіляються члени сукупності. Значення середніх при вивченні різного роду закономірностей географічних явищ, процесів та об'єктів дуже велике. Вони дозволяють: визначити загальну тенденцію розвитку явищ; оцінювати значення окремої величини шляхом порівняння її з середньою; визначати наявність зв'язку між явищами за допомогою аналізу середніх двох або декількох ознак по територіях або часових проміжках.

При обробці даних географічних досліджень як найважливіші характеристики варіаційного ряду застосовуються різноманітні середні значення. Середні величини прийнято розділяти на прості і зважені. Якщо середні значення обчислюються за безпосереднім переліком значень ознаки у кожній одиниці сукупності, то такі середні називаються простими ( незваженими). Якщо середні обчислюються за варіаційним рядом із урахуванням статистичної ваги кожного варіанту, то їх називають зваженими.

Середні величини бувають різного роду. Якщо вид середньої не вказаний, то мається на увазі середня арифметична величина.

Середні величини найчастіше використовують степеневі і структурні (порядкові) середні. Степеневі середні, у свою чергу, підрозділяються на середні арифметичні, середні гармонійні, середні квадратичні, середні кубічні та інші. Структурні середні – це мода, медіана, квартилі, децилі та ін.

Медіаною називається середнє (серединне) значення ознаки ранжированому варіаційного ряду, тобто значення, рівновіддалене від початку та кінця, перебудованого в зростаючому або убуючому порядку.

Модою називається найвірогідніша ( що найчастіше зустрічається) в даному статистичному ряді величина. Мода є найбільшою ординатою кривої розподілу у разі одновіршинного розподілу. У загальному випадку крива розподілу може мати декілька вершин і , відповідно, вона матиме декілька мод.

Вважається, що нормальний розподіл являє собою одну з емпірично перевірених істин і його положення розглядаються як один з фундаментальних законів природи. Точна форма нормального розподілу – «дзвоноподібна крива» – визначається лише двома параметрами: середнім і стандартним відхиленням. Для нормального розподілу характерним є те, що 68% всіх значень знаходяться в межах  $\pm 1$  стандартне відхилення від середнього, а діапазон  $\pm 2$  стандартних відхилення включає 95% значень. Про відхилення від нормального розподілу свідчать коефіцієнти асиметрії та ексцесу. Коефіцієнт асиметрії показує відхилення розподілу від симетричного (нормальний розподіл є абсолютно симетричним, відповідно коефіцієнт асиметрії дорівнює нулю). Додатне значення коефіцієнту асиметрії свідчить про наявність розподілу з «довгим правим хвостом», від'ємне – з «довгим лівим хвостом». Коефіцієнт ексцесу показує «гостроту піку» розподілу (для нормального розподілу коефіцієнт ексцесу також дорівнює нулю). Якщо коефіцієнт ексцесу є додатнім, пік є загостреним, якщо є від'ємним – пік закруглений.



### Практично-семінарське заняття 3:

#### Побудова двовимірних статистичних моделей на матриці вихідних даних

Третім завданням практичної роботи було визначення пар показників, які пов'язані між собою, визначення кількісної оцінки сили зв'язку між ознаками.

Двовимірний наліз – містить дві випадкові величини, утворює двовимірну модель та аналізує дві величини. В ході практичної роботи вивчаються такі двовимірні моделі як: кореляційний та коваріаційний аналізи, які виконується в програмі MINITAB 14. шляхом «Stat» → «Basic Statistics» → «Correlation».

Кореляційний аналіз – метод обробки статистичних даних, що полягає у вивченні тісноти зв'язку між змінними; при цьому порівнюються коефіцієнти кореляції між однією парою або множиною пар ознак для встановлення між ними статистичної взаємодії, що дозволяє виявити найбільш пов'язані змінні та побудувати поверхні їх залежності.

Мета кореляційного аналізу – забезпечити отримання деякої інформації про одну змінну за допомогою іншої змінної. В випадках, коли можливе досягнення мети, говорять, що змінні корелюють. В самому загальному вигляді сприйняття гіпотези про наявність кореляції означає, що зміна значення змінної А відбувається одночасно з пропорційною зміною значення В. Мірою залежності між експериментальними наборами даних є числа – коефіцієнти зв'язку.

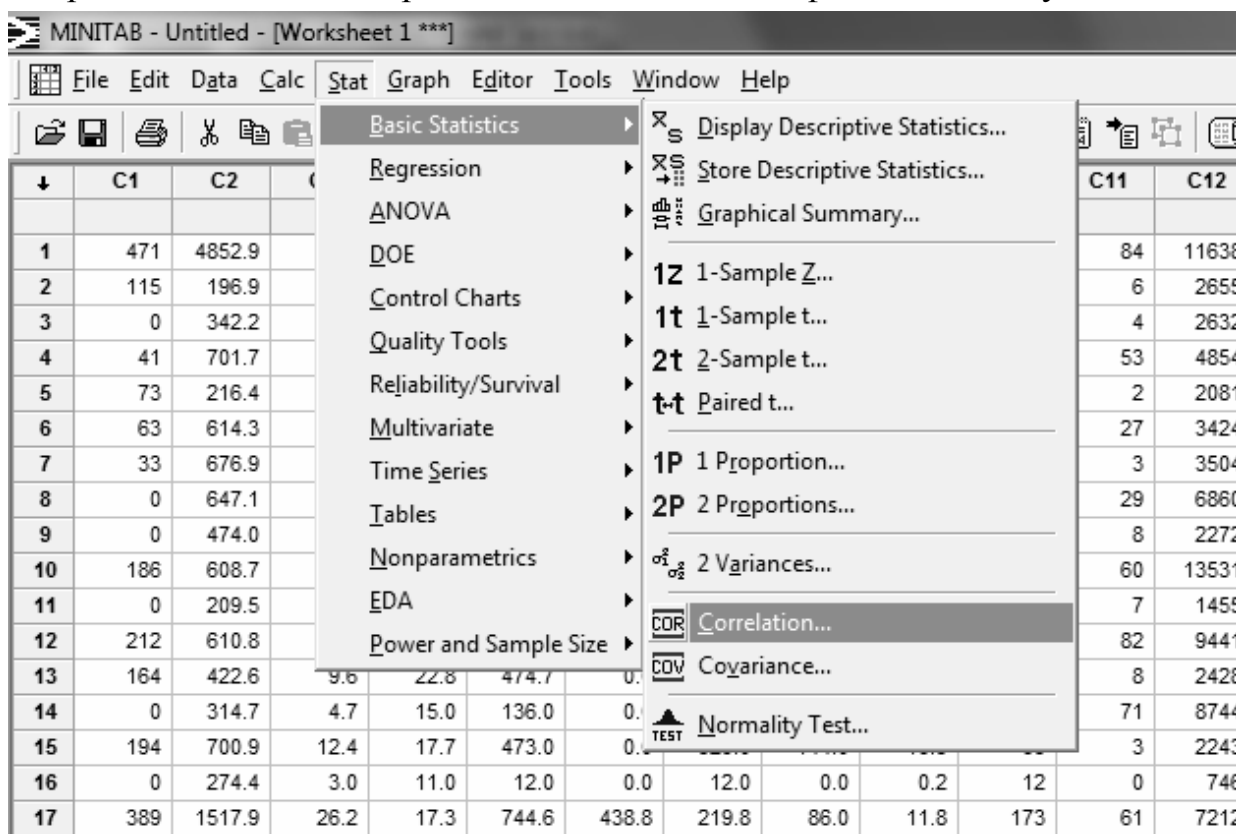


Рис. 3. Алгоритм вибору кореляційного аналізу в програмі MINITAB 1

Головні завдання кореляційного аналізу:

- 1) оцінка за вибірковими даними коефіцієнтів кореляції;
- 2) перевірка значущості вибіркових коефіцієнтів кореляції або кореляційного відношення;
- 3) оцінка близькості виявленого зв'язку до лінійного;
- 4) побудова довірчого інтервалу для коефіцієнтів кореляції.

Визначення сили та напрямку взаємозв'язку між змінними є однією з важливих проблем аналізу даних. В загальному випадку для цього застосовують поняття кореляції. Кореляція є залежністю двох випадкових величин. При цьому, зміна однієї або кількох цих величин призводить до систематичної зміни іншої або інших величин. Математичною мірою кореляції двох випадкових величин слугує коефіцієнт кореляції.

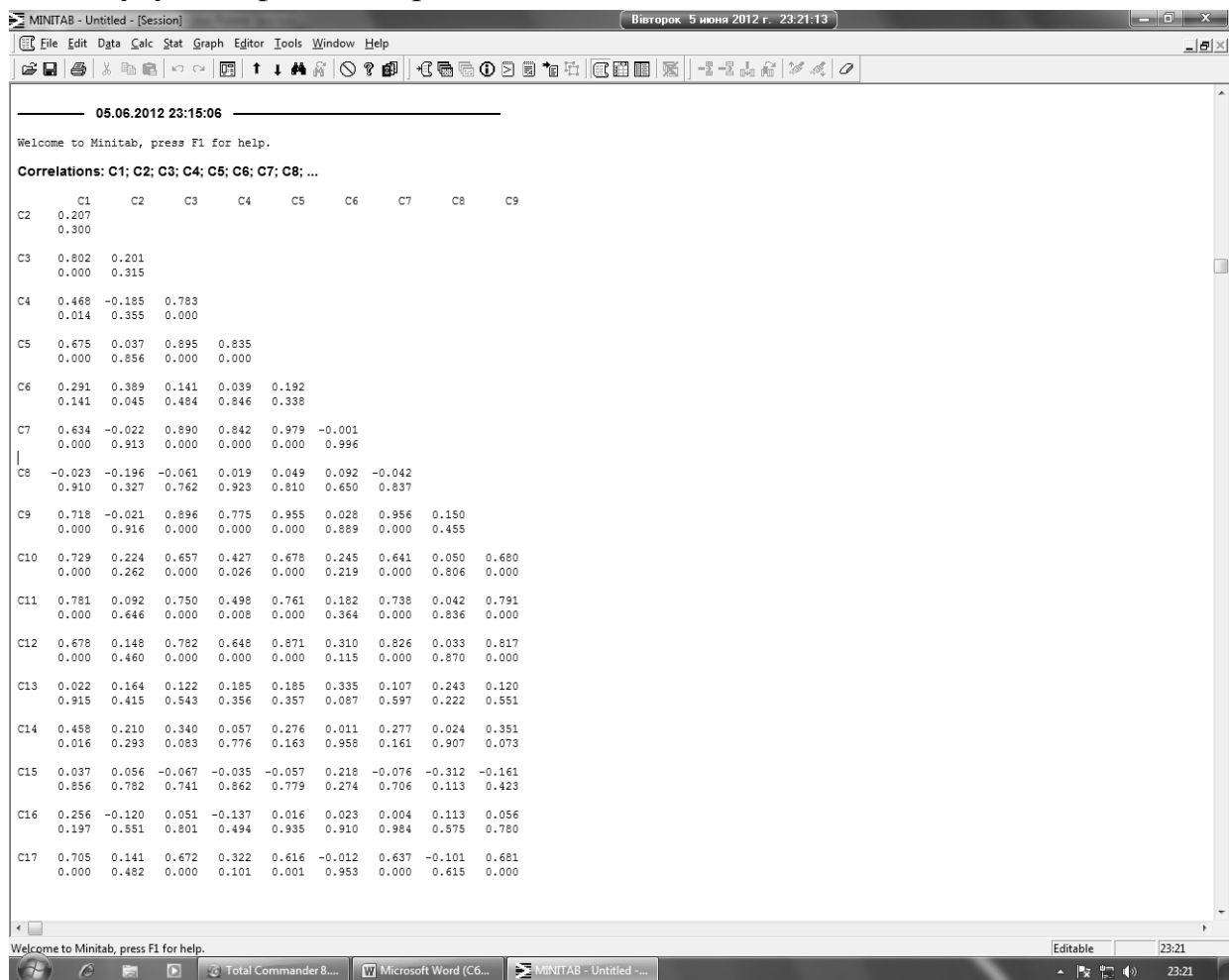


Рис. 4. Кореляційна матриця

Коефіцієнт кореляції приймає значення в межах від -1 до +1. Знак «мінус» означає обернену залежність, «+» – пряму. Якщо коефіцієнт дорівнює нулю, то лінійний зв'язок між динамічними рядами є відсутнім, а якщо одиниці – наявна функціональна залежність. Досить часто помилково вважається, що значення

коефіцієнта кореляції менше 0.2 свідчить про відсутність зв'язку. Насправді він свідчить про наявність дуже слабкого зв'язку, але зв'язок між показниками є.

<i>Коефіцієнт кореляції</i>	<i>Тіснота зв'язку</i>
$> \pm 0.91$	Дуже сильна
$\pm 0.71-0.90$	Сильна
$\pm 0.51-0.70$	
$\pm 0.21-0.50$	Слабка
$< \pm 0.20$	Дуже слабка

Коефіцієнт кореляції, а в загальному випадку кореляційна функція, дозволяють встановити ступінь взаємозв'язку між змінними. Кореляція може бути лінійною або нелінійною в залежності від типу залежності, яка фактично існує між змінними. Досить часто на практиці розглядають тільки лінійну кореляцію (взаємозв'язок), але більш глибокий аналіз потребує використання для дослідження процесів нелінійних залежностей. Складну нелінійну залежність можна спростити, але знати про її існування необхідно для того, щоб побудувати адекватну модель процесу.

У випадку максимальної тісноти зв'язку між показниками на діаграмі розсіювання їх залежність буде представлена прямою лінією. Інколи зустрічаються і з таким явищем як псевдокореляція, тобто кореляція, що обумовлена впливом інших показників, які залишилися поза увагою дослідника. Значимість коефіцієнту кореляції залежить від довжини динамічних рядів. У великих динамічних рядах навіть слабкі залежності будуть значимими, в той час, як у незначних – навіть дуже сильні залежності не є статистично надійними. Тому ведуть мову про надійність кореляційної залежності, яка пов'язана із репрезентативністю вихідних динамічних рядів і свідчить про те, наскільки ймовірно, що виявлена залежність, знову буде виявлена при збільшенні періоду ретроспекції чи екстраполяції на майбутнє. Надійність виявлених залежностей оцінюється за допомогою стандартної статистичної міри –  $r$ -рівня – статистичного рівня значимості. Даний показник знаходиться у оберненій залежності до надійності результату: чим вищим є  $r$ -рівень, тим нижчою є статистична надійність виявленої залежності, і навпаки.

Коваріація – в теорії ймовірностей та математичній статистиці, числова характеристика залежності випадкових величин. Сутність коваріації полягає в тому, що вона виникає внаслідок невизначеності результату перемножування двох сукупностей чисел.

Побудова коваріаційної матриці досить подібна з побудовою матриці кореляції, виконується в програмі MINITAB 14, шляхом «Stat» → «Basic Statistics»

→ «Covariance». В порівнянні з кореляційним зв'язкою коваріаційний не вказує на тісноту зв'язку між певними групами факторів він лише вказує на результат перемножування двох сукупностей чисел.

#### **Практично-семінарське заняття 4:** **Кластер-аналіз та факторний аналіз вихідних даних.** **Інтерпретація отриманих результатів**

Четвертим завданням практичної роботи було визначення територіальної диференціації адміністративно-територіальних одиниць та виявлення значущих факторів суспільно-географічного процесу.

Багатовимірним аналізом називають сукупність різноманітних методів, призначених для вивчення багатовимірних явищ. У багатовимірному просторі досліджувані об'єкти розташовуються, як правило, не рівномірно, а утворюють певні скупчення. Ці скупчення можна розглядати як класи об'єктів. Причому в різних просторах виділяються різні класи. У багатовимірному просторі класифікація більш обґрунтована. Але це означає, що при безмежному зростанні кількості ознак у тій же пропорції зростає точність класифікації.

Існує багато математичних методів і прийомів, які на основі інформації закладеної в матриці даних, дають змогу об'єктивно класифікувати економіко-географічні об'єкти (в тому числі регіоналізувати їх). Такі методи і прийоми об'єднані в багатовимірний аналіз, який розглядається у вузькому і широкому значеннях. У вузькому значенні – це аналіз матриці даних, яка має два і більше стовпчиків, у широкому значенні – це аналіз матриці даних, коли кількість стовпчиків не обмежена знизу. У цьому випадку багатовимірний аналіз включає в себе і одновимірний. Отже одновимірний аналіз можна розглядати як частковий випадок багатовимірного, а багатовимірний – як узагальнення одновимірного. Часто в процесі обробки матриці даних доводиться їх згортати, тобто зводити багатовимірний простір до  $n$ - $k$ -вимірного, а навіть і до одновимірного. В цьому випадку багато ознак замінюються декількома, але такими, кожна з яких синтезує інформацію відповідної групи ознак або є найбільш репрезентативною.

*Кластерний аналіз* передбачає групування об'єктів за подібністю певних характеристик. Термін вперше введений Трайєном /Tryon/, 1939 р. Це один з методів класифікації, що передбачає поділ вихідної сукупності об'єктів на класи (класи, групи). Кластер – це група територіальних одиниць (регіонів), що мають подібні тенденції чи особливості розвитку. З математико-статистичної точки зору, кожний кластер повинен мати такі властивості: густота об'єктів у

межах кластеру має бути вищою за густоту поза ним; можливість відокремлення від інших кластерів тощо. Складність задач кластерного аналізу полягає у тому, що реальні суспільно-географічні об'єкти є багатовимірними, тобто описуються не одним, а певною сукупністю параметрів, тому об'єднання в групи здійснюється у просторі багатьох вимірів. Відповідно критерієм об'єднання регіонів у кластери є мінімум відстані у просторі показників, що їх описують. Звідси – поняття відстані між регіонами у просторі даних.

В ході практичної роботи виконується проведений кластерний аналіз, в програмі MINITAB «Stat» → «Multivariate» → «Cluster Obrevations».

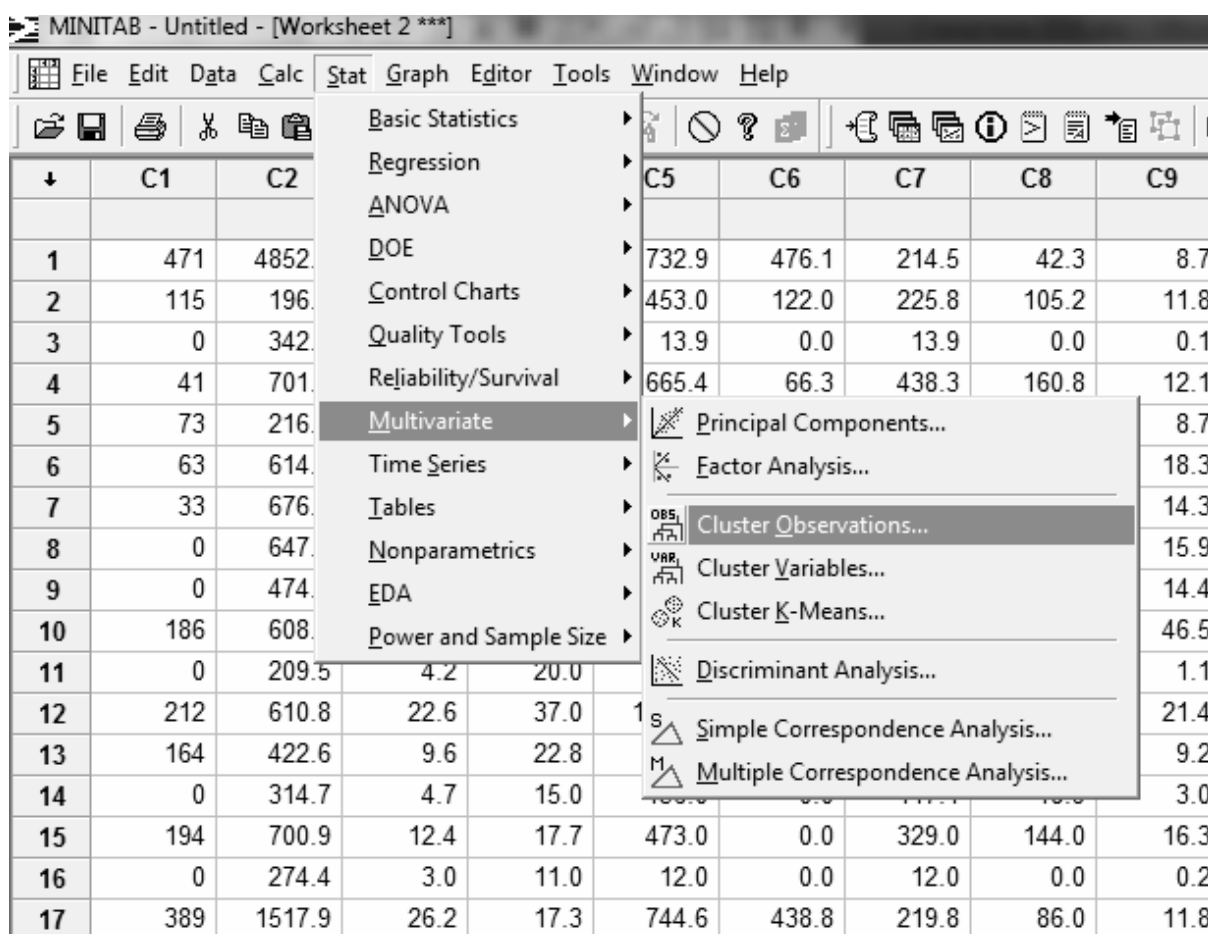


Рис. 5. Алгоритм вибору кластерного аналізу в програмі MINITAB 14

Існують наступні групи методів кластерного аналізу:

- 1) ієрархічні методи;
- 2) ітеративні методи;
- 3) факторні методи;
- 4) методи згущень;
- 5) методи, що використовують теорію графів.

До найпоширенішим в економіці відносять ієрархічні й ітеративні.

Для проведення класифікації необхідно ввести поняття подібності об'єктів за спостережуваними змінними. У кожний кластер (клас, таксон) повинні потрапити об'єкти, що мають подібні характеристики.

Вибір відстані між об'єктами є вузловим моментом дослідження, від нього багато в чому залежить остаточний варіант розбивки об'єктів на класи при даному алгоритмі розбивки.

Алгоритм ієрархічного агломеративного кластерного аналізу можна представити у вигляді послідовності процедур:

- 1) нормуються вихідні дані;
- 2) розраховується матриця відстаней або матриця мір подібності;
- 3) перебуває пара найближчих кластерів, по обраному алгоритмі поєднуються ці два кластери. Новому кластеру привласнюється менший з номерів поєднаних кластерів;
- 4) процедури 2, 3 і 4 повторюються доти, поки всі об'єкти не будуть об'єднані в один кластер або до досягнення заданого "порогу" подібності.

На розрахунок відстаней між об'єктами суттєвий вплив має вибір одиниць виміру показників. Так, наприклад, якщо брати показник обсягів виробництва продукції у мільйонах гривень, різниця між двома територіальними одиницями за цією координатою буде однією, якщо ж у тисячах гривень – у тисячу разів більшою, що впливає на кінцеве групування. Тому кластерний аналіз передбачає здійснення процедури нормалізації даних.

У кластерном аналізі для кількісної оцінки подібності вводиться поняття метрики. Подібність або розходження між класифікованими об'єктами встановлюється залежно від метричної відстані між ними. Якщо кожний об'єкт описується  $k$  ознаками, то він може бути представлений як точка в  $k$ -мірному просторі, і подібність із іншими об'єктами буде визначатися як відповідна відстань. У кластерному аналізі використовуються різні міри відстані між об'єктами:

Найпоширенішими є такі види відстаней:

– евклідова відстань (Euclidean distances), обраховується за теоремою Піфагора: відстані по кожній з координат підносяться до квадрату, а потім із їх суми визначається корінь квадратний;

– манхетенська відстань (відстань міських кварталів, city-block Manhattan distances), обраховується як є сума різниць по кожній з координат;

– відстань Чебишева (Chebychev distance metric), регіони визначають як “різні”, якщо вони різняться по якійсь одній координаті;

– відсоток незгоди (percent disagreement), використовується, коли вихідні дані не мають кількісного виразу.

Методи кластеризації поділяються на дві великі групи – агломеративні (від англ. agglomerate – скупчення) та дивізивні (від англ. division – поділ).

Агломеративні методи передбачають послідовне об'єднання найподібніших регіонів на основі розрахованих відстаней між ними. Процедура кластеризації є такою. На першому кроці, кожний регіон утворює окремий кластер, далі у новий кластер об'єднуються два регіони, міра подібності яких є найбільшою. На останньому кроці всі регіони об'єднуються у один кластер.

Міра подібності кластерів та регіонів визначається такими способами:

- одиничного зв'язку (single linkage, «метод найближчого сусіда»): мінімум найменших відстаней до будь-якого одного регіону у кластері;
- повного зв'язку (complete linkage, «метод найвіддаленіших сусідів»): мінімум найбільших відстаней до будь-якого одного регіону у кластері;
- «середнього» зв'язку: мінімум середньоарифметичного значення відстаней до всіх регіонів у кластері;
- центроїдний мінімум відстаней до центрів ваги кластерів. Центр ваги кластеру визначається як середнє по кожному параметру.

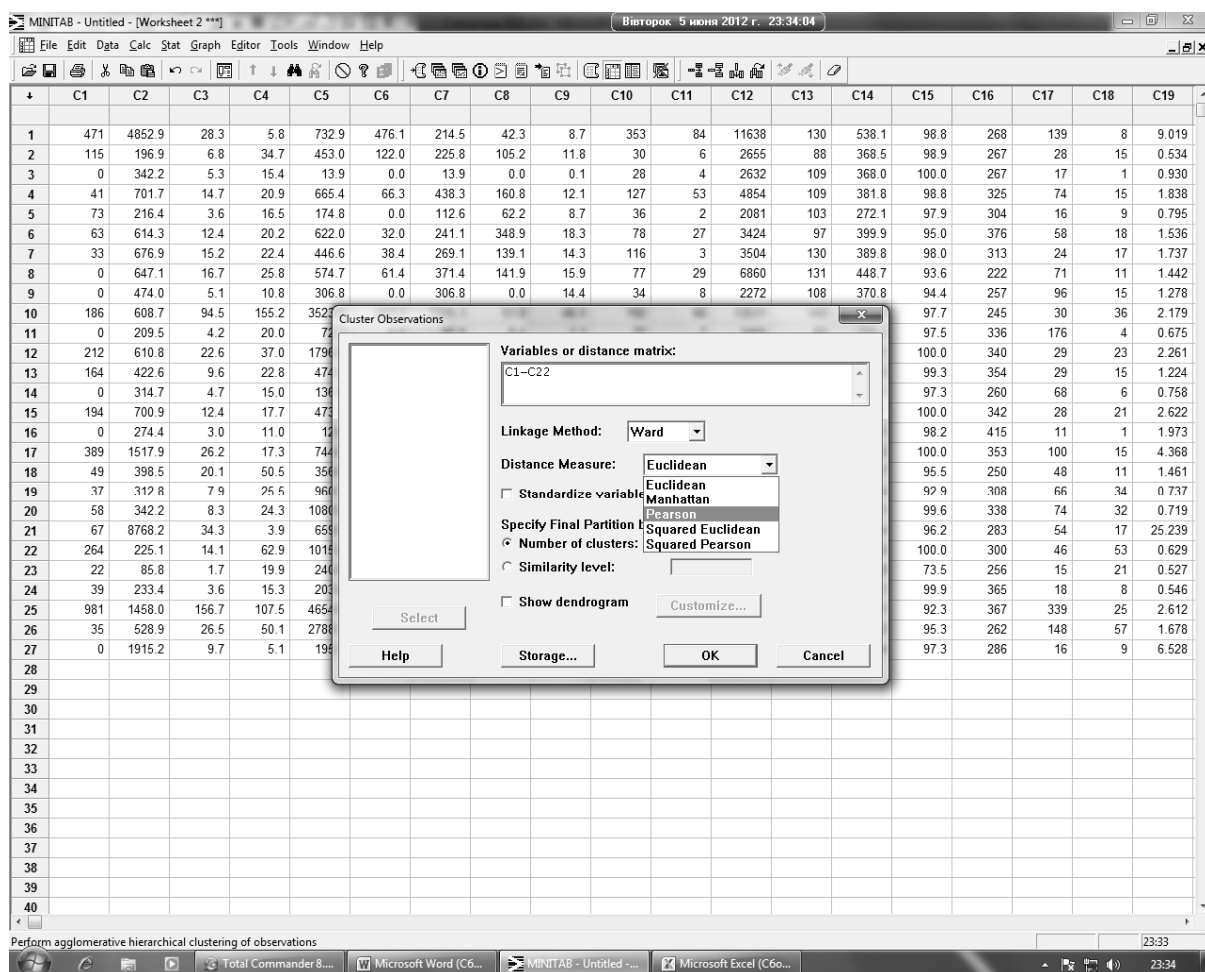


Рис. 6. Вибір відстані при виконанні кластерного аналізу

Результат кластеризації візуалізується у вигляді дендрограми кластеризації (tree diagram, дерево об'єднання), на одній вісі якої відкладаються регіони, на другій – відстані об'єднання (linkage distance).

Дивізивні методи передбачають поетапний (ітераційний) поділ регіонів на задану кількість кластерів. Найпоширенішим серед дивізивних є метод к-середніх (kmeans clustering), що вирішує задачу виділення наперед заданої кількості кластерів, які максимально різняться, тобто знаходяться на найбільших відстанях один від одного.

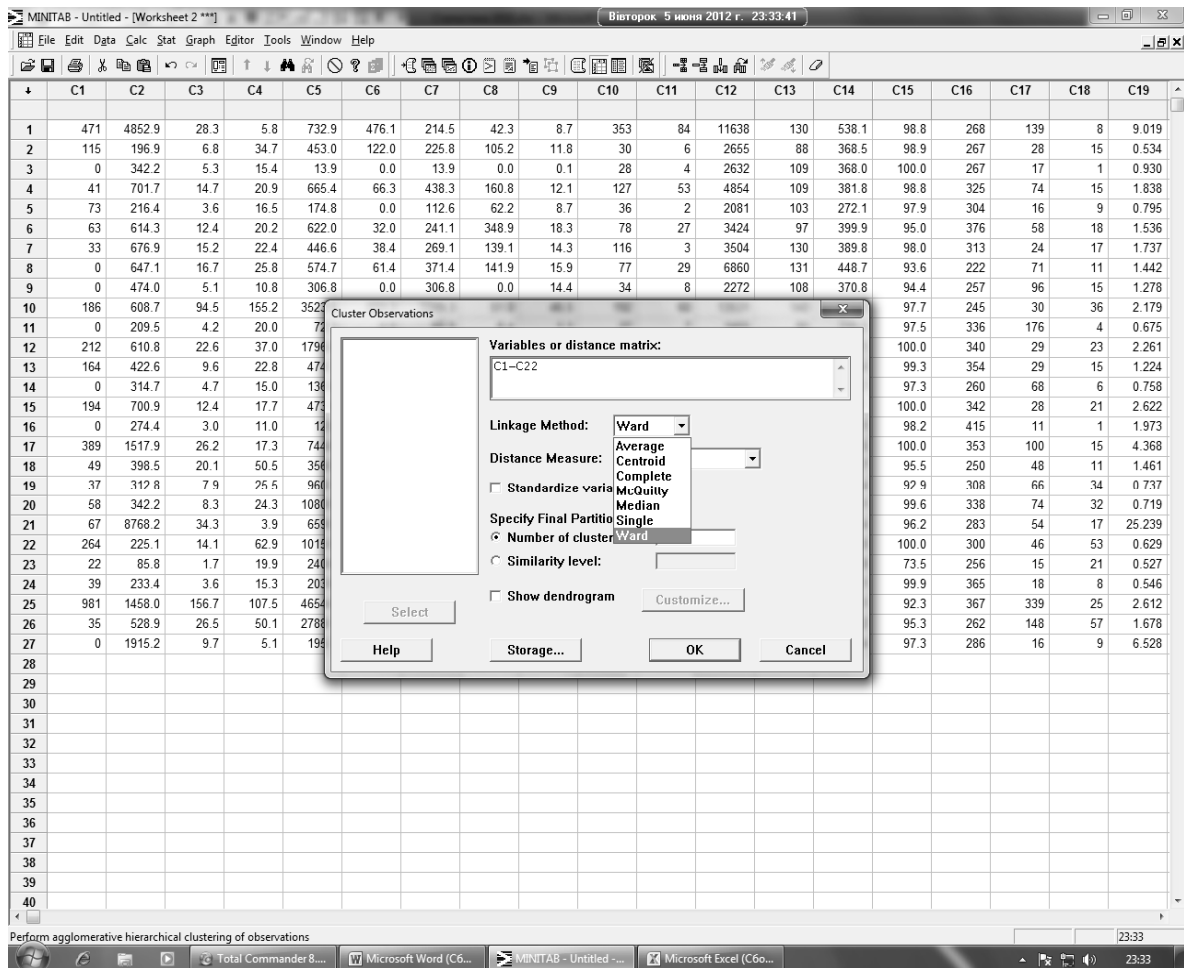


Рис. 7. Вибір методу при виконанні кластерного аналізу

Процедура кластеризації є такою. На першому кроці задається деякий випадковий поділ даних на задану кількість кластерів ( $k$ ), розраховуються центри ваги кластерів. Далі здійснюється переміщення регіонів: кожний регіон відноситься до того кластеру, відстань до центру ваги якого є мінімальною, розраховуються центри ваги нових кластерів. Ця процедура повторюється доки не буде знайдено стабільну конфігурацію, тобто склад кластерів перестане змінюватись.



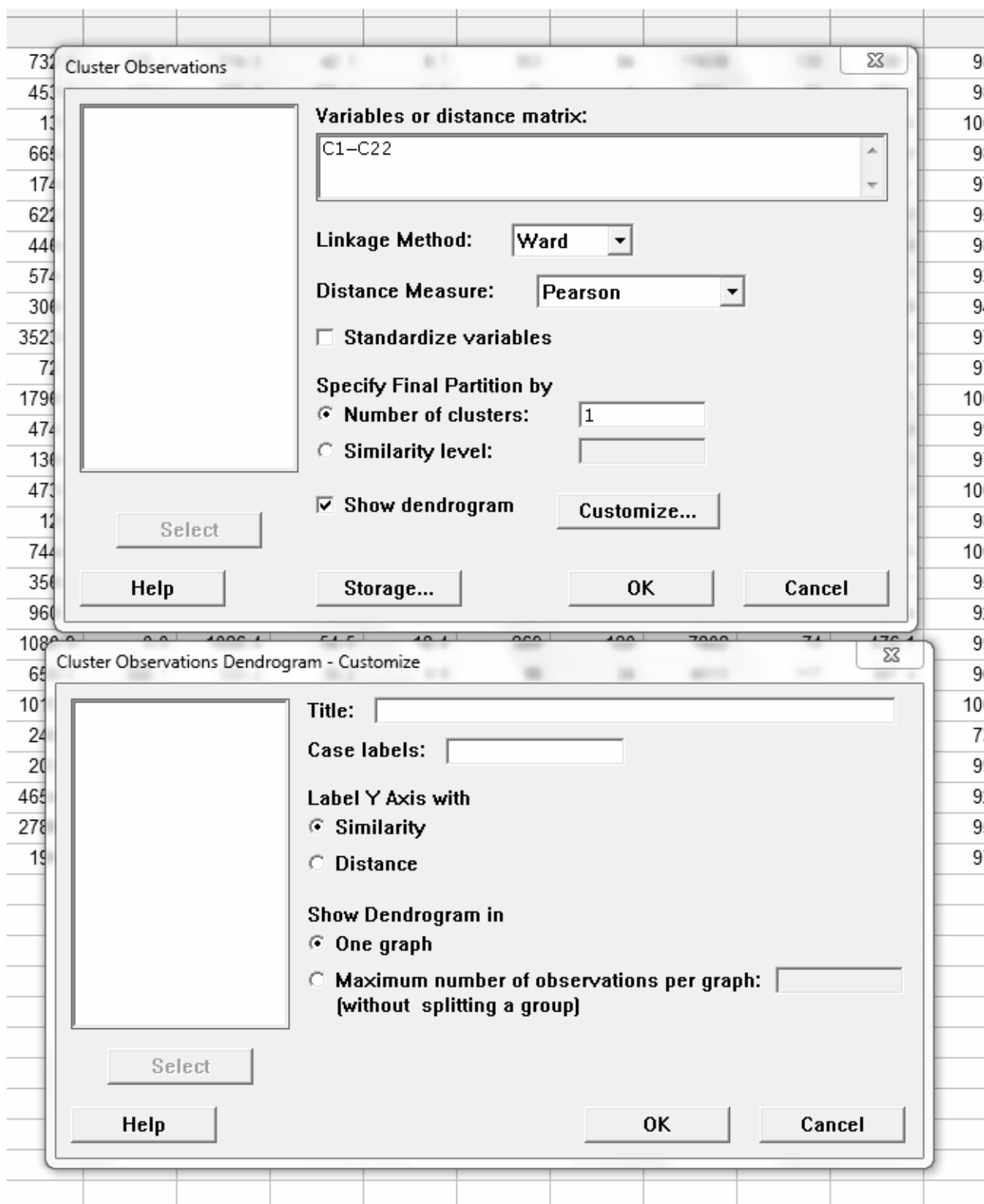


Рис. 8. Вибір основних налаштувань при виконанні кластерного аналізу

Для візуалізації результатів будуються графіки, що являють собою проекції будь-якої пари показників на площину, на яких точки, що відносяться до одного кластеру оконтурюються. Одним з найважливіших та найскладніших питань при кластеризації є вибір оптимальної кількості кластерів. Зазвичай згідно вихі-

дної гіпотези визначається початкова кількість кластерів, а потім змінюючи її, емпіричним шляхом обирають остаточний варіант кластеризації. Для того, щоб оцінити, наскільки виділені кластери різняться між собою, розраховують середні значення базисних показників для кожного кластеру.

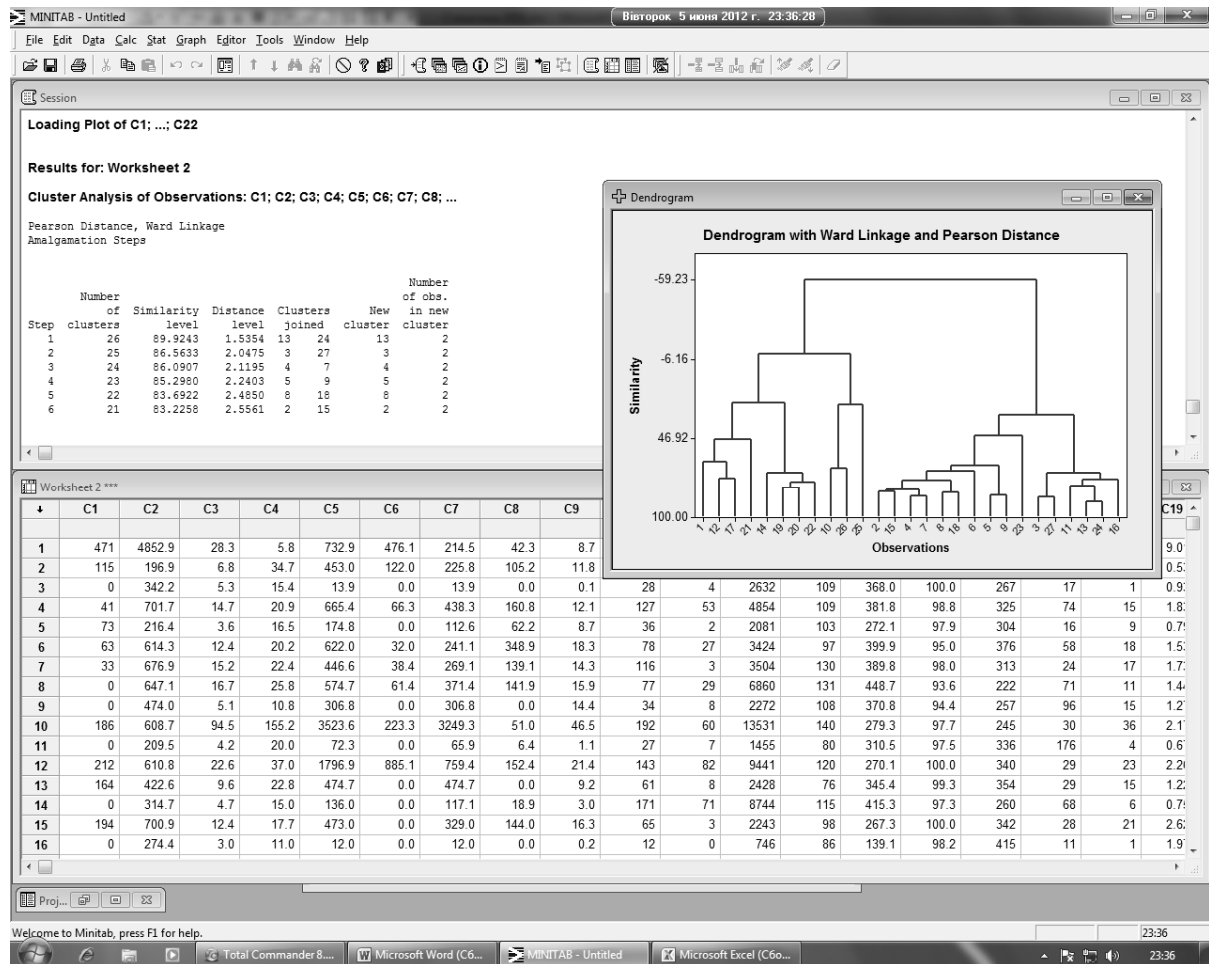


Рис. 9. Приклад результату кластерного аналізу

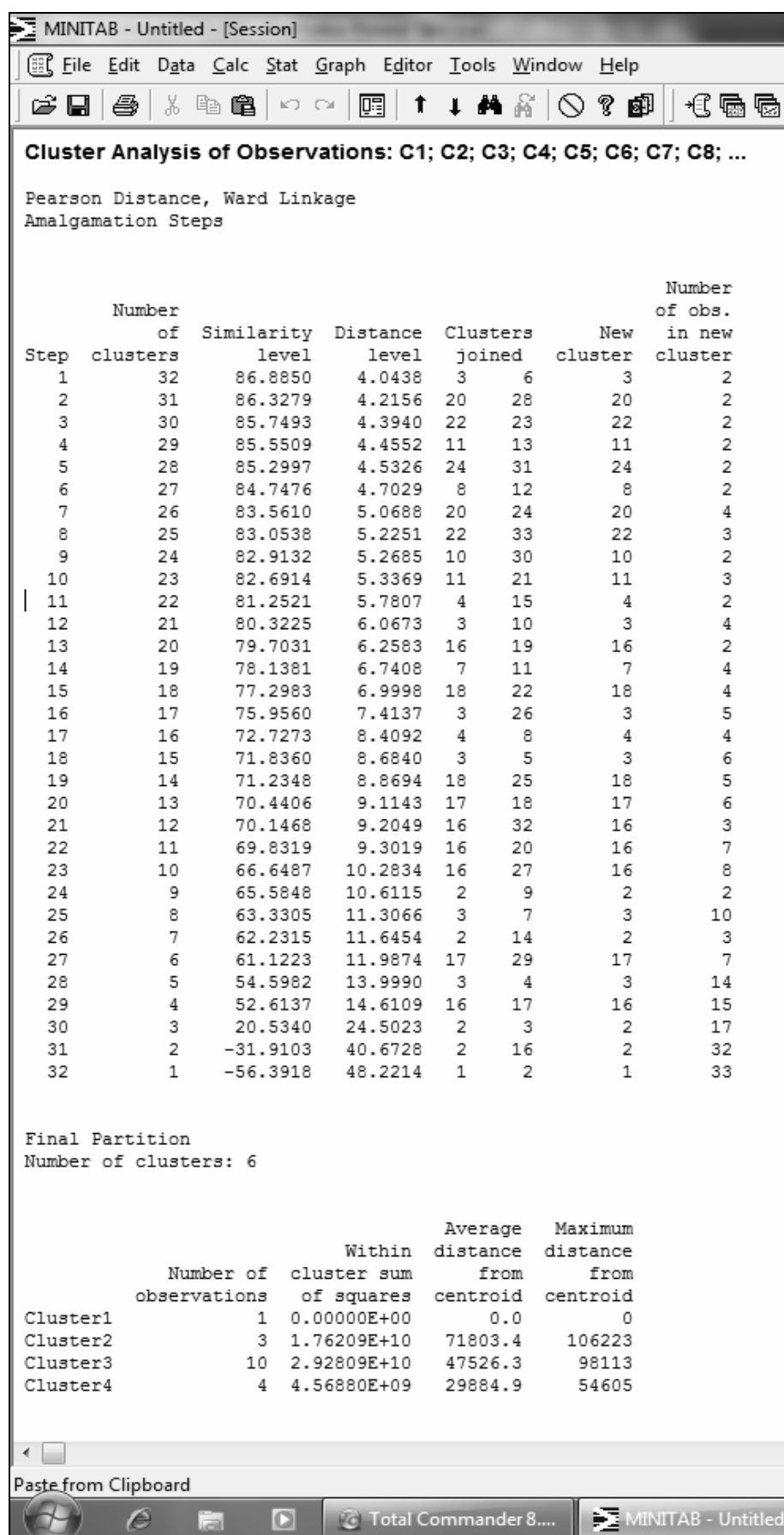


Рис. 10. Приклад протоколу кластерного аналізу

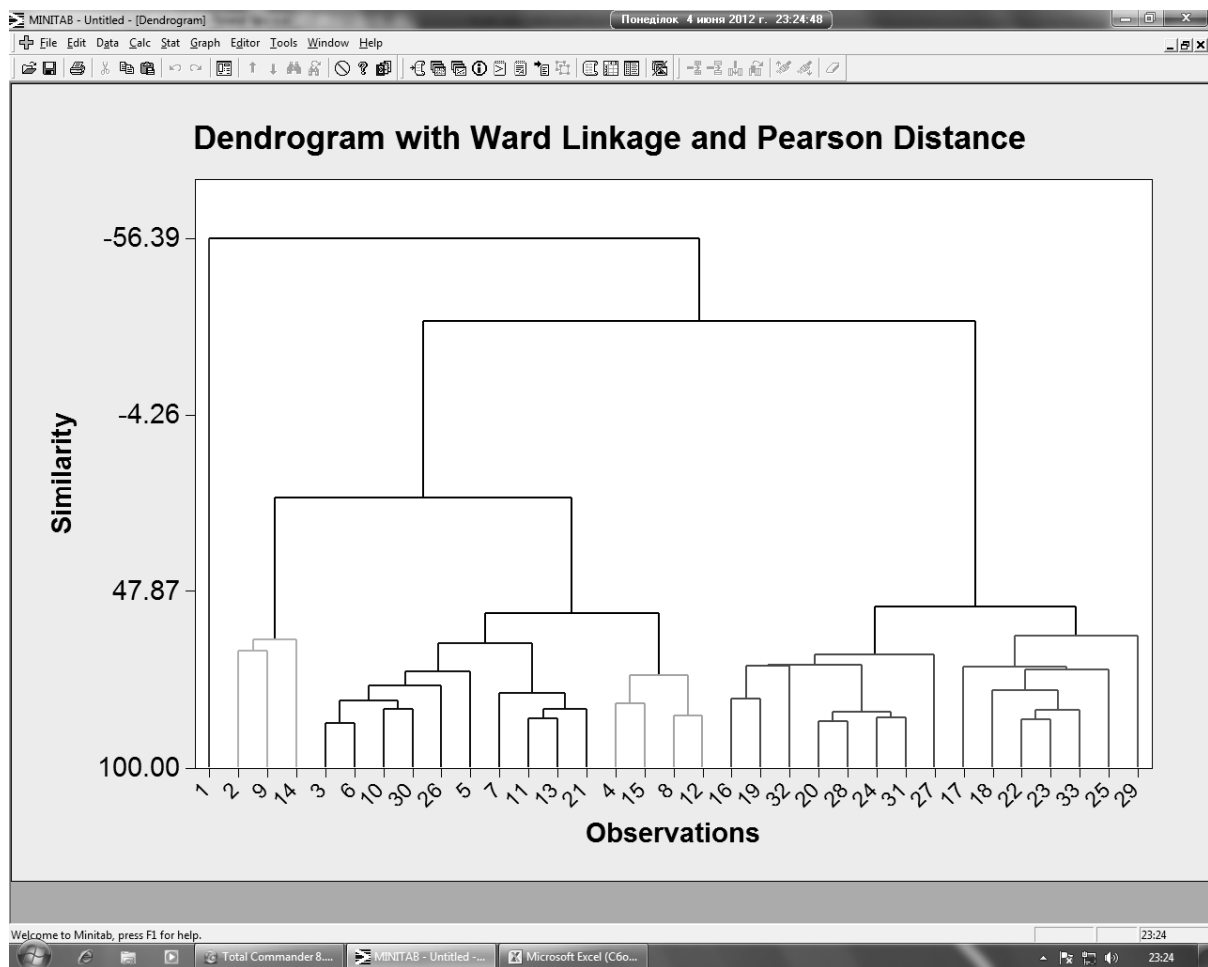


Рис. 11. Приклад дендрограми кластерного аналізу

Кластер-аналіз застосовується при наявності багатовимірних сукупностей статистичних показників, сутність його полягає в об'єднанні їх у групи (кластери) за принципом мінімальної відстані у багатовимірному просторі. Залежно від кількості об'єктів кластеризація та групування виконуються послідовно у декілька кроків таким чином, щоб на останньому кроці в одну загальну групу потрапили усі об'єкти. На перших кроках класифікації формуються найбільш однорідні групи з об'єктів, що мають найбільшу подібність. Поступово, «послаблюючи» критерій відносно подібності об'єктів, об'єднується все більша кількість об'єктів. З кожним кроком до кластерів вищого порядку включаються цілі групи адміністративних районів, які все сильніше різняться між собою. На останньому кроці всі об'єкти об'єднуються в один кластер. Наприкінці процедури отримують групи неоднорідні. В результаті кластерного аналізу отримують багаторівневу ієрархічну класифікацію, яка відображає найбільш суттєві особливості взаємовідношення між об'єктами. Таким чином, отримані кластери – це група територіальних одиниць, що мають подібні особливості розвитку. Такий аналіз проводиться для територіальних об'єктів за рядом показників, для подальшого групування районів та виявлення стійких груп.

Факторний аналіз – це багатовимірний статистичний метод, який застосовується для вивчення взаємозв’язків між значеннями змінних. Він дозволяє визначити групи чинників впливу та дослідити значення кожної змінної у дії фактора у різний час.

В ході практичної роботи виконується факторний аналіз методом головних компонент, в MINITAB 14 шляхом «Stat» → «Multivariate» → «Principal Components». Факторний аналіз дає змогу оцінити чинники розвитку.

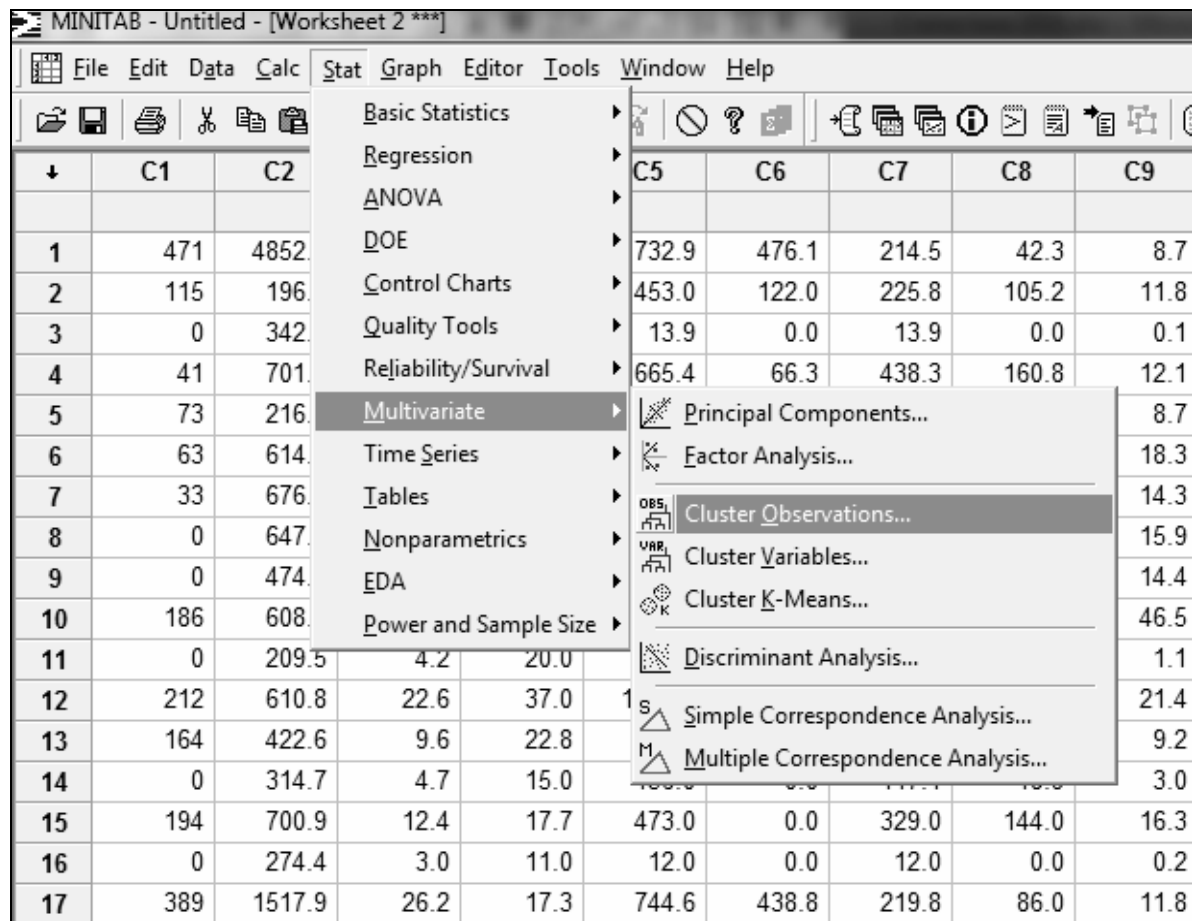


Рис. 12. Алгоритм вибору факторного аналізу в програмі MINITAB 14

В основі побудови моделей факторного аналізу лежить твердження про те, що множину взаємопов’язаних показників, які характеризують певний процес, можна представити меншою кількістю гіпотетичних змінних – факторів та множиною незалежних залишків. Інформаційною базою розрахунків служать просторові ряди показників демографічного розвитку у розрізі адміністративно-територіальних одиниць регіону за базові роки. Набір показників є суб’єктивним; основним принципом вибору показників були достовірність та співставність. Після обрання оптимальної кількості факторів, виконується інтерпретація результатів, враховуючи показники, які сформували виявлені фактори.

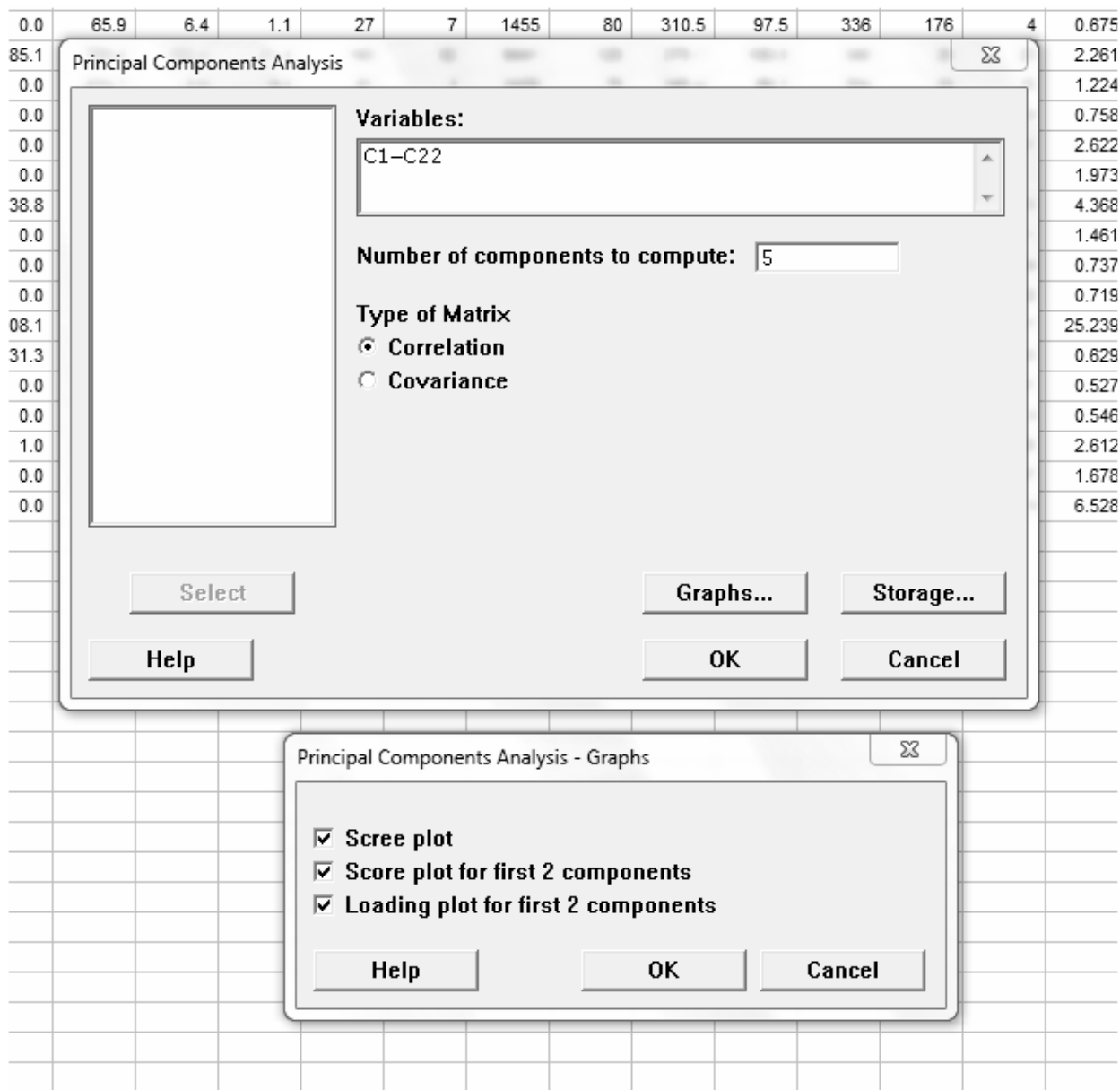


Рис. 13. Вибір вихідних даних та елементів виконання факторного аналізу

Зазвичай фактори, що впливають на розвиток суспільно-просторових процесів, характеризуються не одним, а множиною показників, що мають між собою певний зв'язок (корелюють). Тому виникає потреба у їх об'єднанні у певну кількість груп за подібністю впливу, в результаті чого визначаються на перший погляд "приховані" (латентні) фактори, кількість яких, зрозуміло, є меншою за кількість вихідних показників. Тобто фактори ідентифікуються.

Застосування факторного аналізу передбачає:

- по-перше, скорочення кількості показників (змінних), що на мові математичної статистики зветься редукцією даних;
- по-друге, визначення структури взаємозв'язків між показниками (змінними), тобто класифікацію показників, інтерпретацію факторів.

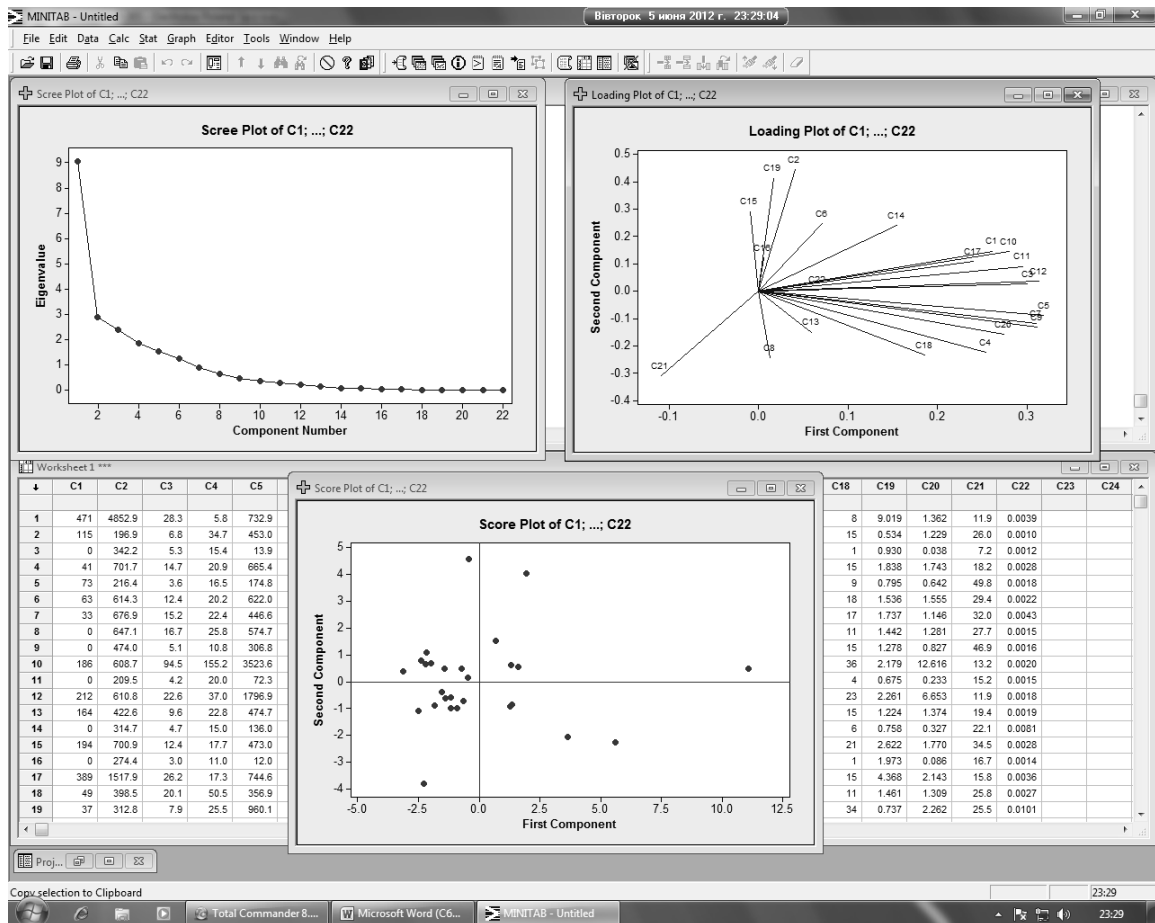


Рис. 14. Приклад результату факторного аналізу

Таким чином, факторний аналіз використовується і як метод зменшення кількості даних, і як метод класифікації одночасно. В основі побудови моделей факторного аналізу лежить твердження про те, що множину взаємозв'язаних показників, які характеризують певний процес, можна представити меншою кількістю гіпотетичних змінних – факторів та множиною незалежних залишків. Зміст факторного аналізу полягає у лінійному перетворенні  $n$ -вимірного простору у  $k$ -вимірний. Іншими словами, за допомогою факторного аналізу систему  $n$  показників можна замінити значно меншою кількістю ( $k$ ) факторів.

Факторні навантаження знаходяться в межах від -1 до +1. Знак “+” чи “-” вказує на наявність прямої або оберненої залежності між показником і фактором. Зміст факторів визначають показники (змінні), що мають найбільші факторні навантаження (найближчі по модулю до одиниці).

Процедура факторного аналізує включає такі складові:

– визначення кількості факторів. На першому кроці розрахунків кількість факторів обрано рівною кількості показників та обраховано абсолютні, відносні та кумулятивні значення дисперсії кожного з факторів. Визначають, яка кількість факторів є оптимальною. Зазвичай користуються одним з трьох критеріїв:

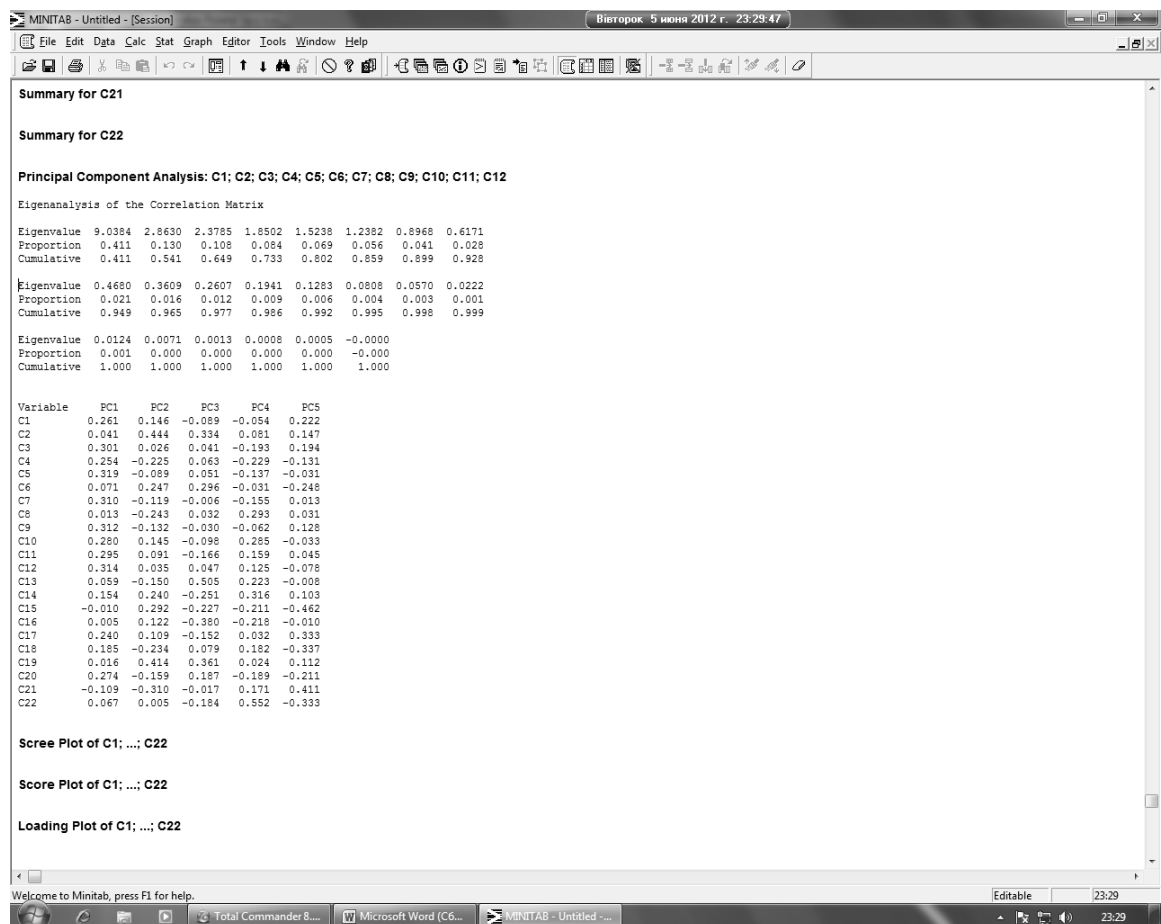


Рис. 15. Приклад протоколу факторного аналізу

– за критерієм Кайзера (Kaiser, 1960): обираються тільки фактори із дисперсією більше 1. Враховується те, що, якщо фактор не виділяє дисперсію, еквівалентну, хоча б дисперсії одного показника, то він відкидається. У наведеному прикладі варто виділяти два фактори;

– за кумулятивним відсотком: в якості визначальних обираються ті фактори, що сумарно охоплюють приблизно три чверті вихідної інформації (кумулятивний відсоток має перевищувати 75%). У наведеному прикладі два перших фактори пояснюють 90,5% загальної дисперсії – виділяємо два фактори;

– за критерієм “кам’янистого осипу” Кеттеля (Cattell, 1966): на графіку дисперсій (plot of eigenvalue, scree plot) знаходиться таке місце, де зменшення дисперсії зліва направо максимально уповільнюється. Передбачається, що справа від цієї точки знаходиться лише “факторіальний осип” (термін “осип” запозичений з геології, де означає уламки гірських порід, що накопичуються в нижній частині скелястого схилу). Відповідно до цього критерію у наведеному прикладі виділяємо 2 або 3 фактори: обернення осей координат (factor rotation), ідентифікація та інтерпретація факторів. У результаті обернення отримується остаточна матриця факторних навантажень:



– обрахунок факторних ваг. Факторні ваги (factor scores) – це показники, що відіграють роль оцінок вкладів територіальних одиниць у кожний з факторів. Матриця факторних ваг обчислюється шляхом множення матриці вихідних даних на матрицю факторних навантажень. Вони трактуються як відносні оцінки вияву певного фактору і служать основою для їх групування;

– аналіз тенденцій.

Останнім завданням практичної роботи є інтерпретація отриманих (в ході виконання всіх видів аналізу) результатів.

## **ПИТАННЯ ДО ПІДСУМКОВОГО МОДУЛЬНОГО КОНТРОЛЮ**

1. Об'єкт та мета курсу.
2. Визначення математичних методів та моделювання.
3. Поняття «модель». Класифікації моделей: натурні, аналогові, математичні.
4. Історія розвитку і використання моделей в природознавстві і суспільних науках.
5. Три рівні математизації суспільної географії: рівень параметризації географічних явищ та об'єктів, рівень емпіричних моделей, рівень теоретичних моделей.
6. Місце методів ідеалізації, формалізації, математичних методів, методів моделювання в методології суспільно-географічного дослідження.
7. Протиріччя системного і синергетичного підходів.
8. Перспективи розвитку моделювання в суспільній географії.
9. Специфіка виникнення, функціонування і розвитку природних та соціальних об'єктів.
10. Багатофакторність суспільно-географічних процесів.
11. Складність інтерпретації причинно-наслідкових зв'язків в соціогеосистемах.
12. Організація географічного середовища. Локальні та інтегральні процеси.
13. Поняття про ймовірність.
14. Принципово випадковий характер суспільно-географічних процесів.
15. Особливості і методи отримання суспільно-географічної інформації.
16. Одновимірні статистичні моделі.
17. Поняття про випадкову величину, дискретні і неперервні випадкові величини.
18. Числові характеристики випадкових величин: частота, частість, функція розподілу.
19. Оцінки центру розподілу: математичне очікування, медіана, мода.

20. Оцінки розкиду випадкової величини: центральні моменти другого, третього і четвертого порядків.
21. Основні закони одновимірного розподілу.
22. Вибірковий метод, вимоги до вибірових сукупностей.
23. Поняття про надійну ймовірність.
24. Точкова та інтервальна оцінка статистик.
25. Вимоги до оцінок параметрів суспільно-географічних об'єктів.
26. Поняття про перевірку статистичних гіпотез.
27. Умови застосування одновимірних статистичних моделей.
28. Двовимірні статистичні моделі.
29. Поняття про двовимірну випадкову величину.
30. Умови формування двовимірних випадкових величин у суспільно-географічних дослідженнях.
31. Особливості формування вибірових сукупностей.
32. Особливості інтерпретації зв'язків компонентів двовимірних випадкових величин.
33. Поняття про форму залежності і силу зв'язку двох випадкових величин.
34. Функціональні і статистичні залежності.
35. Двовимірний регресійний аналіз, поняття про умовний розподіл, лінійні, нелінійні рівняння регресії, обчислення регресійних коефіцієнтів.
36. Побудова довірчого інтервалу для рівнянь регресії.
37. Кореляційний аналіз, поняття про коефіцієнт кореляції, кореляційне відношення.
38. Умови коректності кореляційного аналізу.
39. Нелінійні перетворення випадкових величин для приведення до нормально-го закону розподілу.
40. Особливості використання регресійно-кореляційного аналізу у суспільно-географічних дослідженнях.
41. Багатовимірні статистичні моделі.
42. Поняття про багатовимірну випадкову величину.
43. Поняття про матрицю вихідних даних.
44. Багатовимірний кореляційний аналіз, парний, частковий, множинний коефіцієнт кореляції. Особливості використання багатовимірного кореляційного аналізу.
45. Багатовимірний регресійний аналіз, визначення значущості часткових коефіцієнтів регресії. Особливості застосування багатовимірного регресійного аналізу.
46. Поняття про багатовимірний ознаковий простір, визначення відстані у багатовимірному просторі.

47. Використання теорії графів для класифікації суспільно-географічних об'єктів.
48. Кластер-аналіз, визначення дистанційних коефіцієнтів, принципи утворення кластерів.
49. Задачі розпізнання образів, поняття про власну область об'єкту, вирішальну функцію, вирішальне правило. Алгоритм розпізнавання образів.
50. Дискримінантний аналіз у задачах класифікації.
51. Факторний аналіз, його різновиди, інтерпретація результатів.
52. Поняття про просторову змінну, ознаки і властивості просторових змінних.
53. Відмінності моделей просторових змінних від статистичних моделей.
54. Суспільно-географічні об'єкти як генератори полів просторових змінних.
55. Мінливість та анізотропність полів просторових змінних.
56. Закономірна та випадкова складові мінливості, що спостерігається.
57. Поняття про фон та аномалію.
58. Поняття про тренд.
59. Тренд-аналіз як метод розділення фону та аномалії.
60. Задачі тренд-аналізу в суспільній географії.
61. Загальна характеристика методів згладжування випадкових полів.
62. Метод ковзного статистичного вікна.
63. Принципи розрахунку локального середнього.
64. Розмір ковзного статистичного вікна як параметр генералізації карти апроксимації випадкового поля.
65. Аналіз залишків та виділення аномалій.
66. Інтерпретація результатів тренд-аналізу.
67. Загальний огляд методів апроксимації поверхні тренда єдиною функцією координат простору.
68. Принципи оцінки точності апроксимації поверхні тренду.
69. Моделі алгебраїчних поліномів, вибір порядку поліному.
70. Моделі тригонометричних поліномів.
71. Використання диференційних рівнянь для апроксимації поверхні тренду.

### **Приклад підсумкового модульного контролю**

**I рівень – репродуктивно-фактажний (максимум 6 балів)**

*Оберіть одну вірну відповідь; за кожну вірну відповідь –1 бал*

1. Твердження стосовно поняття «математичні методи»:

*а) використання апарату математичної або формальної логіки для вирішення конкретних задач, наприклад, для кількісного аналізу даних;*

- б) деякий матеріал чи подумки представлений об'єкт або явище, що є спрощеною версією модельованого об'єкта або явища (прототипа) і в достатній мірі повторює властивості, суттєві для цілей конкретного моделювання;
  - в) методи дослідження явищ і процесів, що ґрунтуються на заміні конкретного об'єкта досліджень (оригіналу) іншим, подібним до нього;
  - г) наукові методи опису та вивчення масових явищ, що допускають кількісний вираз.
2. Яке значення з наведених нижче може приймати імовірність (p):
- а)  $p = -0,5$
  - б)  $p = 0,6$
  - в)  $p = 1,5$
  - г)  $p = 3,2$ .
3. Метод, заснований на втраті інформації, яка відбувається в результаті поділу об'єктів на кластери і яка вимірюється за допомогою загальної суми квадратів відхилень кожного об'єкта від центру свого кластеру – це:
- а) близьколежачий сусід
  - б) Варда.
  - в) зв'язок між групами
  - г) медіанна кластеризація
4. До погано організованої системи можна віднести:
- а) суспільство
  - б) прилади
  - в) фізичні процеси
  - г) хімічні процеси
5. За яким підходом імовірність не є результатом нашої недосконалості, а є об'єктивним атрибутом матерії:
- а) гносеологічним
  - б) географічним
  - в) епістемологічним
  - г) онтологічним.
6. Класифікація процесів, їх аналіз та оцінка інформації здійснюється на етапі математичного моделювання, який має назву ...
- а) змістовний аналіз
  - б) кількісний аналіз.
  - в) спостереження-опис-вимірювання
  - г) прогноз-практика

## **II Рівень – Поняттєвий (максимум 10 балів)**

*Дайте визначення поняттю, за кожену вірну відповідь – 2 бали*

Статистичні методи – це

---

---

---

Частота – це

---

---

---

Мода – це

---

---

---

Дисперсія – це

---

---

---

Регресійний аналіз – це

---

---

---

## **III Рівень – Теоретичний (максимум 24 бали)**

*Дайте розгорнуту відповідь на запитання:*

1. Природні та соціальні об'єкти: специфіка виникнення, функціонування і розвитку (8 балів).
2. Лінійний дискримінантний аналіз і його застосування в суспільній географії (8 балів).
3. Кластер-аналіз і його модифікації (8 балів).

## ОЦІНЮВАННЯ РЕЗУЛЬТАТІВ НАВЧАЛЬНИХ ДОСЯГНЕНЬ СТУДЕНТІВ

	Модуль 1		Модуль 2		Іспит	Сума балів
	Поточний модульний контроль	Практична робота	Поточний модульний контроль	Практична робота		
Розподіл балів	20	10	20	10	40	100
Сума балів за модулем	30		30			

## ШКАЛА ОЦІНЮВАННЯ

Сума балів за всі види навчальної діяльності протягом семестру	Оцінка ECTS	Оцінка за національною шкалою	
		для екзамену, курсової роботи (проекту), практики	для заліку
90–100	A	відмінно	зараховано
80–89	B	добре	
70–79	C		
60–69	D	задовільно	
50–59	E		
1–49	FX	незадовільно	не зараховано

## КОРОТКИЙ ГЛОСАРІЙ

<b>Абсолютні статистичні величини</b>	Первинний вид статистичних показників, які характеризують чисельність сукупності, розмір (обсяг, рівень) досліджуваного явища і його чинників у конкретних межах часу і місця. Абсолютні величини завжди мають назву і можуть бути виражені в певних одиницях виміру – натуральних (км, кг, шт.), умовних (умовне паливо), грошових (грн., \$), комплексних (кВт/год., т/км) та часових (рік, сек.) тощо.
<b>Аналітичне вирівнювання (згладжування)</b>	Ефективний метод виявлення і кількісної оцінки тенденції зміни випадкової величини в ряду спостережень. При цьому фактичні рівні динамічного ряду замінюють теоретичними, розрахованими на основі рівняння регресії. На основі фактичних даних ряду динаміки підбирається математична функція – «трендова крива» (лінійна, параболічна, показова, експоненційна та ін.), за допомогою якої описується основна тенденція. Час розглядається як незалежна змінна, а рівні ряду виступають як функція цієї змінної $\hat{f}(t)$ .
<b>Анізотропія поля</b>	Проявляється у залежності значення параметру поля від напрямку, у якому визначається. Коефіцієнт анізотропії є відношенням найбільшого і найменшого значення параметру в одній точці, але визначеного в різних напрямках.
<b>Апроксимація емпіричних залежностей, полів</b>	Наближене представлення емпіричної залежності, як функції часу або простору, певною математичною функцією або рівнянням регресії (див. аналітичне вирівнювання). Поля ознак об'єктів апроксимуються поверхнями як єдиними функціями координат простору або побудованими за принципом локальних середніх.
<b>Асиметричний розподіл</b>	Розподіл ознак, у якому частоти по обидва боки від центра змінюються неоднаково, тобто вершина розподілу зміщена. Існує правостороння ( $x_{\text{сер}} > M_e > M_o$ ) і лівостороння ( $x_{\text{сер}} < M_e < M_o$ ) асиметрії. Асиметрія виникає як результат обмеженої варіації ознак в одному напрямку чи впливу домінуючої причини розвитку явища, яка призводить до зміщення центра розподілу.
<b>Атрибутивний ряд</b>	Послідовність статистичних показників, яку утворюють за якісною ознакою. Варіанти розташовують у їх логічній послідовності.
<b>Багатовимірна статистична модель</b>	Модель, яка містить більше двох випадкових величин. Використовується для встановлення характеристик взаємозв'язку між кількома ознаками об'єкту дослідження.
<b>Багатовимірний аналіз</b>	Сукупність різноманітних методів, призначених для вивчення явищ (процесів, об'єктів) у багатовимірному ознаковому просторі.

<b>Багатовимірний ознаковий простір</b>	Віртуальний простір, у якому базисом є не звичайні лінійні характеристики об'єктів (ширина, довжина, висота), а кількісні показники (параметри) ознак об'єктів. Вимірність простору не обмежується. Місцеположення кожного об'єкту у багатовимірному просторі фіксується точкою з координатами, які відповідають параметрам об'єкту.
<b>Варіаційний ряд</b>	Послідовність будь-яких чисел, розташованих в порядку зростання (зменшення) їх величин. Варіаційні ряди будуються на основі кількісної групуючої ознаки, складаються з варіантів (окремих значень варіаційної ознаки, що їх приймає ця ознака в ряді розподілу) та частот (чисельностей окремих варіант або кожної з груп варіаційного ряду).
<b>Варіація</b>	Зміна значень конкретної ознаки при переході від однієї одиниці сукупності до іншої, це окреме значення ознаки кожного елемента (одиниці) сукупності.
<b>Вибіркова сукупність (вибірка)</b>	Сукупність одиниць спостереження, відібраних із генеральної сукупності за певними правилами, які забезпечують репрезентативність вибірки. Вибірка повинна бути статистично стійкою і репрезентативною. Методи обробки вибірових даних залежать від її обсягу. Існує статистика малих (до 10 елементів), середніх (10 – 60 елементів) і великих (більше 60 елементів) вибірових сукупностей. Надійність результатів дослідження вибірових даних зростає із збільшенням обсягу вибірки.
<b>Вибірковий метод</b>	Основний метод математичної статистики, який полягає у формуванні вибіркової сукупності шляхом відбору частини елементів цілого (генеральної сукупності), яким притаманні властивості усієї спільноти. Знаходячи статистичні характеристики вибіркової сукупності, їх узагальнюють за певними правилами на всю генеральну сукупність. Вибірка репрезентативна (представницька) тоді, коли структура досліджуваних ознак об'єкту в даній вибірці незначним чином відрізняється від структури ознак в генеральній сукупності.
<b>Випадкова величина</b>	Величина, яка приймає в результаті досліду одне з багатьох значень, при цьому поява того чи іншого значення цієї величини до його визначення (вимірювання) принципово неможливо точно передбачити. Випадкові величини, які є результатом вимірювання і можуть приймати безкінечне число значень при послідовному визначенні називаються неперервними. Дискретні випадкові величини мають обмежене фіксоване число можливих значень (підкидання монети, кубика тощо).



<b>Відносні статистичні величини</b>	Похідні показники, які характеризують кількісне співвідношення статистичних даних. Їх обчислюють як частку від ділення двох абсолютних величин. Чисельником є порівнювана (поточна) величина, що вивчається, а знаменником – база (основа) порівняння, величина з якою порівнюють і яка є своєрідним вимірювачем, еталоном, оптимальним рівнем. Співставляти можна одновимірні показники, які відносяться до різних періодів, різних об'єктів, територій, а також різновимірні.
<b>Генеральна сукупність</b>	Нескінченне або кінцеве (досить велике) число елементів або компонентів, що складаються з якісно однорідних показників; це абсолютно всі можливі результати певної події, повна сукупність точок, замірів тощо, які можна отримати на даному об'єкті.
<b>Геоінформаційний метод</b>	Метод дослідження, який набуває дедалі більшого значення і поширення. Полягає у створенні геоінформаційних систем (ГІС), які є сучасним засобом збирання, збереження та аналізу різноманітних відомостей про територію. Використання ГІС-технологій суттєво розширює можливості аналізу.
<b>Геоінформаційні (геопросторові) дані</b>	Сукупність статистичних показників, яка мають чітку просторову прив'язку – набір координат, що однозначно визначають місцезположення досліджуваного об'єкту у реальному фізичному чи віртуальному просторі.
<b>Геоінформаційні системи (ГІС)</b>	Інформаційна система, що забезпечує збір, збереження, відображення, поширення, аналіз і перетворення геопросторових даних. ГІС містить дані про географічні об'єкти у цифровому представленні (растрові, векторні тощо) і набір функцій, що реалізують геоінформаційні технології (візуалізація, картографічне моделювання, просторовий аналіз, проектування тощо).
<b>Географічний простір</b>	Сукупність відношень між географічними об'єктами, які розташовані на певній території (геоторії) та які розвиваються з часом.
<b>Геосистема</b>	Матеріальна система, у якій відношення між елементами опосередковані територією (геоторією). Характеризується цілісністю, взаємозв'язками і взаємодією елементів, структурною стійкістю та емерджентністю.
<b>Гіперповерхня</b>	Узагальнення поняття звичайної поверхні 3-вимірного простору на випадок n-вимірного простору. Зазвичай гіперповерхня задається одним рівнянням зв'язку між координатами. Якщо в евклідовому n-вимірному просторі гіперповерхня задається лінійним рівнянням, то вона називається гіперплощиною.

<b>Гістограма</b>	Сходинковий графік для інтервального варіаційного ряду. Утворені прямокутники пропорційні за висотою частотам варіантів для кожного інтервалу. У випадку нерівних інтервалів висота прямокутників пропорційна щільності розподілу ознаки у конкретному інтервалі. Гістограму можна перетворити на полігон, з'єднавши середини вершин стовпчиків лінією.
<b>Двовимірна статистична модель</b>	Модель, що містить дві випадкові величини і в залежності від їх типу може бути неперервною, дискретною або змішаною. Використовується для встановлення характеристик взаємозв'язку між двома ознаками об'єкту дослідження.
<b>Двовимірний аналіз</b>	Аналіз двовимірної випадкової величини, який дозволяє встановити наявність залежності між двома випадковими величинами, її форму, тісноту, знайти рівняння для її опису. Основні види аналізу – кореляційний і регресійний.
<b>Дендрограма</b>	Графічне зображення у двовимірній проекції результатів кластер-аналізу, тобто, групування об'єктів за їх подібністю.
<b>Детермінована модель</b>	Модель, функціонування якої визначене через відомі однозначні відношення станів і подій, і в якій один і той же заданий вхід буде завжди видавати строго один результат (наприклад, модель відомої хімічної реакції).
<b>Детерміновані системи</b>	Системи, процеси в яких взаємозв'язані так, що можна відслідкувати ланцюг причинно-наслідкових залежностей у функціональному вигляді (строга відповідність функції аргументу). Детермінізм тісно пов'язаний із ступенем організації системи.
<b>Динамічний (часовий) ряд</b>	Сукупність значень статистичних показників, які розташовані у хронологічному порядку. Подається у вигляді таблиці чи графіка (на осі абсцис відкладають шкалу часу, на ординаті – шкалу рівнів ряду).
<b>Дискримінантний аналіз</b>	Різновид багатовимірної аналізу, призначеного для класифікації об'єктів за сукупністю даних.
<b>Дисперсія</b>	Оцінка ступеня відхилення значень випадкової величини від центру розподілу. Більші значення дисперсії свідчать про більший розкид значень випадкової величини від центру розподілу.
<b>Довірча ймовірність</b>	Ймовірність накриття довірчим інтервалом істинного значення параметра. Довжина довірчого інтервалу пропорційна величині довірчої ймовірності. Найбільш вживані стандартні значення довірчої ймовірності наступні: $P = 0.9; 0.95; 0.99; 0.999$ .

<b>Екстраполяція</b>	Знаходження невідомих рівнів (значень) за межами наявних фактичних даних динамічного ряду за визначеними аналітичними рівняннями тренда (математичними моделями).
<b>Закон розподілу ймовірностей</b>	Основне поняття теорії ймовірностей, яке для випадкової величини визначає множину можливих подій з ймовірностями їхнього набуття. Інакше, співвідношення, що встановлює зв'язок між можливими значеннями випадкової величини і відповідними ймовірностями. Характеризується функцією щільності розподілу (відображає ймовірність потрапляння емпіричної випадкової величини в заданий інтервал) та інтегральною функцією розподілу (відображає ймовірність набуття емпіричною випадковою величиною значення більше або менше заданого).
<b>Ймовірність</b>	Числова характеристика можливості того, що випадкова подія відбудеться в умовах, які можуть бути відтворені необмежену кількість разів.
<b>Інтервальна оцінка</b>	Відображається на числовій вісі довірчим інтервалом, в якому із заданою ймовірністю знаходиться параметр, що оцінюється. Верхня і нижня межі довірчого інтервалу визначаються величиною можливого відхилення вибіркової оцінки параметра від його істинного значення і називаються відповідно максимальним і мінімальним гарантованими значеннями параметру із заданою ймовірністю.
<b>Інтерполяція</b>	Знаходження показника (приблизний розрахунок) в середині ряду на основі закономірності розвитку явища за період, що досліджується. Точність такого розрахунку залежить від стабільності показників динаміки – абсолютних приростів, темпів росту тощо. Обчислення проводять на основі середнього абсолютного приросту або середнього темпу росту.
<b>Інформаційно-суспільно-географічний простір</b>	Простір антропосфери, де існують соціум, господарство і природні системи, територіальна організація, властивості і зв'язок (між собою та зовнішнім середовищем), яких зумовлені складною взаємодією полів різних ознак, а єдність і оптимізація здійснюється через інтерференцію та взаємодію їхніх інформаційних полів.
<b>Інформація</b>	Існує кілька груп визначень інформації: – будь-які відомості, дані, що передаються різними каналами; – будь-які відомості, дані, що зменшують ступінь невизначеності про навколишнє середовище; – міра неоднорідності, розмаїття, впорядкованості, структурованості, організованості систем (атрибутивна концепція інформації);

	– запам'ятований вибір в точці біфуркації одного варіанта розвитку з декількох можливих (рівноправних).
<b>Кластерний аналіз</b>	Вид багатовимірною аналізу, завданням якого є розбиття заданої вибірки об'єктів (ситуацій) на підмножини, що називаються кластерами, так, щоб кожен кластер складався з схожих об'єктів, а об'єкти різних кластерів істотно відрізнялися.
<b>Коваріація</b>	Числова характеристика взаємозв'язку (взаємовпливу) випадкових величин. Сутність коваріації полягає в тому, що вона відображає результат стохастичної взаємодії географічних об'єктів під впливом великої кількості факторів.
<b>Коефіцієнт асиметрії</b>	Кількісна оцінка асиметричності (скошеності) графіку функції щільності розподілу випадкової величини.
<b>Коефіцієнт ексцесу</b>	Міра гостровершинності графіку функції щільності розподілу випадкової величини.
<b>Коефіцієнт кореляції</b>	Визначається за результатами кореляційного аналізу. Парний коефіцієнт кореляції характеризує тісноту лінійного зв'язку між двома випадковими величинами, множинний коефіцієнт кореляції характеризує тісноту залежності однієї випадкової величини від множини інших випадкових величин. Частковий коефіцієнт кореляції відображає тісноту зв'язку між двома випадковими величинами з врахуванням впливу множини інших випадкових величин. Коефіцієнт кореляції змінюється в межах від -1 до 1, зростання його значення по модулю свідчить про посилення зв'язку, при прямому зв'язку між випадковими величинами коефіцієнта кореляції більше 0, при зворотному – менше 0. Коефіцієнт кореляції можна розглядати як показник тісноти зв'язку тільки при виконанні наступних умов: а) залежність між випадковими величинами лінійна; б) розподіл випадкових величин відповідає нормальному закону.
<b>Кореляційний аналіз</b>	Метод обробки статистичних даних, за допомогою якого вимірюється тіснота зв'язку між випадковими величинами
<b>Кумулята і огіва</b>	Криві накопичених (кумулятивних) частот або часток, їх використовують для зображення як дискретних так і інтервальних рядів. Будуючи кумуляту, на абсцисі відкладають варіанти, на ординаті – нагромаджені частоти. У разі побудови огіви, яка є дзеркальним відображенням кумуляти, навпаки. За цими кривими визначають, скільки одиниць сукупності, або яка їх частка не перевищує певного значення групувальної ознаки(для дискретного ряду) чи верхньої межі відповідного інтервалу (для інтервального ряду).

<b>Математична модель</b>	У загальному випадку система рівнянь, які описують досліджуваний процес або явище.
<b>Математичне очікування</b>	Одна з основних числових характеристик кожної випадкової величини. Є узагальненим поняттям середнього значення сукупності чисел, коли елементи множини значень цієї сукупності мають різну «вагу», ціну, важливість, пріоритет, що характерно для випадкових величин.
<b>Математичні методи</b>	Узагальнення використання апарату математичної або формальної логіки для вирішення конкретних задач, наприклад, для кількісного аналізу даних.
<b>Медіана</b>	Величина, що знаходиться рівно в середині ряду значень випадкової величини, розташованих у зростаючому або спадаючому порядку. Медіана поділяє ряд значень ознаки на дві рівні частини, по обидві частини від неї розміщується однакова кількість одиниць сукупності: 50% одиниць досліджуваної сукупності матимуть значення ознаки менше, ніж медіана, а 50% – більше.
<b>Метод плинної (ковзної) середньої</b>	Метод дослідження, при якому послідовно знаходять середні арифметичні значення (локальні середні) з певних $m$ рівнів ряду (інтервалу осереднення), поступово зміщуючись на один рівень по ряду на кожному кроці. Знайдене значення локальної середньої стосується того періоду, який знаходиться в середині інтервалу осереднення. Метод дає можливість згладжувати емпіричні дані, встановлювати тенденції динамічного ряду.
<b>Метод апроксимації поверхні тренда</b>	Опис тенденцій просторової змінності показників досліджуваних об'єктів у вигляді статистичної поверхні. Для цього застосовуються дві групи методів: локальної середньої (ковзне статистичне вікно) та представлення поверхні як єдиної функції координат простору.
<b>Метод згладжування випадкових полів</b>	Полягає у застосуванні ковзного статистичного вікна, у якому послідовно знаходяться локальні середні (частіше з урахуванням відстані від опорних точок до центра вікна) з поступовим зміщенням вікна по просторовим координатам. В результаті отримують множину точок з локальними середніми значеннями поля, що утворюють регулярну сітку. Далі використовуються різні методи візуалізації згладжуваних поверхонь.
<b>Мода</b>	Значення випадкової величини, що найчастіше зустрічається в сукупності спостережень, тобто має найбільшу ймовірність (вагу, частоту, частку). Модальне значення відповідає піку (найбільшому значенню ймовірності) графіку функції щільності ймовірності розподілу випадкової величини.

<b>Моделювання</b>	Метод дослідження явищ і процесів, що ґрунтується на заміні конкретного об'єкту досліджень (оригіналу) іншим, подібним до нього (моделлю) відповідно до певних правил, що встановлені в теорії подібності та аналогії. Дослідження моделі дає нову інформацію про оригінал.
<b>Модель</b>	Деякий матеріальний чи подумки представлений об'єкт або явище, що є спрощеною версією модельованого об'єкта або явища (оригіналу) і в достатній мірі відображає властивості, суттєві для цілей конкретного моделювання (опускаючи несуттєві властивості, в яких він може відрізнятися від оригіналу). Це матеріально реалізована або уявно представлена система, яка, відображаючи об'єкт дослідження (оригінал), об'єктивно здатна замінити його так, що її вивчення дає нову інформацію про оригінал.
<b>Неоднорідність (мінливість) поля</b>	Полягає у тому, що значення параметру поля залежить від координат точки і тому змінюється у просторі.
<b>Нормальний розподіл (розподіл Гауса)</b>	Розподіл випадкової величини, який відображається дзвоноподібною кривою (симетрична відносно тх ординати) щільності розподілу. Такий розподіл є результатом впливу на значення ознаки чи властивості об'єкту необмеженої кількості незалежних один від одного факторів з невеликою силою впливу, як це буває в природі. Нормальний розподіл повністю визначений двома параметрами – математичним очікуванням (середньою арифметичною) та стандартним (середнім квадратичним) відхиленням. Значення ознаки при нормальному розподілі переважно зосереджуються біля центру розподілу.
<b>Одновимірна статистична модель</b>	Статистична модель, яка містить одну випадкову величину. В основі її аналізу є обчислення характеристик варіаційного ряду: математичного очікування, дисперсії, стандартного відхилення, коефіцієнтів варіації, асиметрії, ексцесу та інших статистик.
<b>Полігон частот</b>	Відображення згрупованого або дискретного варіаційного ряду у вигляді ламаної лінії в прямокутній системі координат. Значення ознаки (центри інтервалів групування) відкладається на осі абсцис (X), а частоти (частки, щільність) – на осі ординат (Y).
<b>Прогнозування</b>	Процес передбачення майбутнього стану предмета чи явища на основі аналізу історії його розвитку, систематична інформація про якісні й кількісні характеристики розвитку явища в перспективі (наприклад, прогноз погоди тощо). Результатом прогнозування є прогноз — знання про майбутнє і про ймовірний розвиток сьогочасних тенденцій.

<b>Регресійний аналіз</b>	Статистичний метод дослідження залежності між залежною змінною $Y_i$ однією або кількома незалежними змінними $X_1, X_2, \dots, X_p$ . В першому наближенні застосовують лінійний регресійний аналіз (рівняння лінії, площини, гіперплощини). В більш складних випадках використовуються нелінійні рівняння (параболи, гіперболи, поліноми різних ступенів тощо).
<b>Симетричний розподіл</b>	Розподіл, в якому частоти будь-яких двох варіантів, рівновіддалених в обидва боки від центра, рівні між собою.
<b>Синергетична парадигма</b>	Виникла на основі синергетики – науки про взаємодію і саморозвиток систем (Г. Хакен) наприкінці ХХ ст. Передбачає всебічний аналіз об'єкту дослідження (міждисциплінарний комплексний підхід), врахування якомога більшої кількості діючих факторів, орієнтує на нелінійне представлення процесів. Синергетична парадигма є продовженням системної парадигми.
<b>Система</b>	Сукупність взаємопов'язаних та взаємодіючих елементів (підсистем), яка має наступні ознаки: цілісність, стійкість структури, емерджентність.
<b>Системний аналіз</b>	Велика група методів для дослідження систем і їх взаємодії. Основні напрями: функціонально-системний та структурно-системний аналіз. Функціонально-системний аналіз головним чином спрямований на дослідження зовнішньої адаптації системи, тобто, її реакції на зміни зовнішнього середовища і взаємодії з ним, іншими системами рівного порядку, рідше – між власними підсистемами. Структурно-системний аналіз має головним завданням дослідження внутрішньої адаптації системи, перш за все – структурних зв'язків між її елементами, умов гомеостазу, структурної стійкості тощо.
<b>Соціально-географічна система (соціогеосистема)</b>	Гетерогенна система, що вміщує різні за рівнем узагальнення та ієрархії соціальні елементи або підсистеми, а також техногенні абіотичні і біогенні елементи (підсистеми), що знаходяться у взаємодії через потоки речовини, енергії та інформації в географічному просторово-часовому континуумі.
<b>Соціально-географічний процес (соціогеопроцес)</b>	Послідовна закономірна зміна ситуацій у розвитку різних соціумів в історичному і географічному контексті, як зміна соціальних складових соціогеосистем у просторово-часовому континуумі.
<b>Статистика</b>	Поняття багатозначне, у науковій і практичній сферах під ним розуміють: – вид інформації, тобто зібрані статистичні дані (підсумкові цифрові показники) про різноманітні масові явища суспільного життя;

	<p>– форму практичної діяльності, спрямовану на збирання, упорядкування, оброблення та інтерпретацію даних (системи статистичних установ, дослідницькі центри тощо);</p> <p>– суспільну науку – спеціальну наукову дисципліну, яка займається розробкою теоретичних положень і методів, що використовуються статистичною практикою, а також навчальну дисципліну, яку викладають у вищих закладах освіти всіх рівнів для підготовки фахівців найвищої кваліфікації;</p> <p>«міру», «оцінку» або «характеристику», тобто певні статистичні показники, наприклад, узагальнені показники вибіркової сукупності (середня, частка) тощо.</p>
<b>Статистична закономірність</b>	Кількісне представлення закономірностей змін у просторі та часі масових явищ і процесів суспільного життя, які складаються з множини елементів (одиниць сукупності). Це форма прояву причинного зв'язку, виражена у послідовності, регулярності, повторюваності подій, якщо причини (умови), що викликали ці події не змінюються або мають незначні зміни.
<b>Статистична сукупність</b>	Множина значень параметру досліджуваного об'єкту (однорідних у певному відношенні), поєднаних спільними властивостями, умовами та причинами існування і розвитку.
<b>Статистичний показник</b>	Кількісна або напівкількісна характеристика властивості (ознаки) об'єкту досліджень, яка певним чином відображає механізми процесів чи стан об'єкту. Кількісно визначені показники називаються параметрами. Статистичні показники поділяються на абсолютні, відносні та середні величини і утворюють статистичні сукупності.
<b>Статистичний ряд розподілу</b>	Впорядкований розподіл одиниць досліджуваної сукупності на групи за групувальною (варіативною) ознакою. Вони характеризують склад (структуру) досліджуваного явища, дозволяють робити висновки про однорідність сукупності, межі її зміни, закономірності розвитку досліджуваного об'єкта. Залежно від ознаки статистичні ряди розподілу діляться на: атрибутивні (якісні) та варіаційні (кількісні), останні поділяються на дискретні та інтервальні
<b>Статистичні методи</b>	Набір методів аналізу статистичних даних, наприклад, в соціально – економічній географії про господарство і населення регіонів та країн для з'ясування особливостей просторової взаємодії різних територіально-економічних систем. У цьому аспекті важливу роль відіграє статистика як ефективний інструмент дослідження економічного і соціального стану конкретної території.



<b>Стохастичний (статистичний) зв'язок</b>	Вид зв'язку між випадковими змінними, для якого характерно, що кожному значенню факторної (незалежної) ознаки відповідає певна множина (спектр) значень результативної (залежної) ознаки. Тобто причинна залежність проявляється не в кожному окремому випадку, а в загальному, при великій кількості спостережень. На відміну від функціональних, стохастичні зв'язки неоднозначні і характеризуються ймовірністю. Частіше всього такі зв'язки виявляються як узгодженість варіації двох чи більше ознак.
<b>Суспільно-географічний процес</b>	Послідовна і закономірна зміна суспільно-географічних явищ. Є одним з основних понять соціально-економічної географії і узагальнює всі процеси, що протікають у суспільстві і техносфері (антропосфері).
<b>Точкова оцінка</b>	Відповідає точці на числовій вісі і є випадковою величиною. Повинна відповідати наступним вимогам: переконливість, незміщеність, максимальна ефективність. Точкова оцінка не містить інформацію про точність отриманого результату.
<b>Тренд-аналіз</b>	В найбільш загальному вигляді тренд-аналіз є методом дослідження певних просторово-часових закономірностей у розвитку суспільно-географічного процесу, відображених у наборі статистичних сукупностей параметрів об'єкту дослідження. Включає кілька груп методів аналізу емпіричних (статистичних) даних для виділення закономірної (фонові) і випадкової (аномальної) складових поля параметру певної ознаки об'єкту дослідження. Відповідно до цього зазвичай будь-яке поле параметру (яке розглядається як сукупність випадкових величин) розкладається на закономірну (трендову), аномальну (випадкову) складові і так званий залишок – різницю між ними. У такому випадку говорять про зняття тренду і сукупність залишку розглядають як окремий об'єкт дослідження.
<b>Факторний аналіз</b>	Велика група статистичних методів аналізу багатовимірних статистичних сукупностей, для яких загальним завданням є виявлення латентних (прихованих) гіпотетичних факторів впливу на досліджуваний процес або об'єкт. Як правило ці фактори мають певну змістовну інтерпретацію, що більш повно розкриває особливості взаємодії окремих факторів (чинників), відображених в статистичних показниках.
<b>Фон та аномалію</b>	З точки зору математичної статистики фон є певним середнім (нормативним) рівнем значень поля будь-якого параметру, визначеним з урахуванням довірчого інтервалу. Значення поля, що виходять за межі довірчого інтервалу,

	розглядаються як випадкові відхилення і складають аномалії. Дуже часто виділення аномалій є одним із основних завдань дослідження і досягається застосуванням тренд-аналізу.
<b>Формалізація</b>	Представлення об'єкту дослідження у вигляді, придатному для аналізу засобами формальної (математичної) логіки. Полягає у перетворенні якісних (концептуальних) уявлень про структуру, властивості, особливості динаміки об'єкту у формальний опис об'єкту у вигляді однозначно визначених математичних залежностей.
<b>Функціональний зв'язок</b>	Характеризується повною відповідністю між зміною незалежної ознаки (аргументу) і результатної величини (функції). При функціональному зв'язку кожному можливо-му значенню факторної ознаки (x) відповідає одне і тільки одне чітко визначене значення результативної ознаки (y). Завдяки цьому функціональну залежність можна описати математичними формулами. Такий зв'язок маємо у фізичних, хімічних процесах – залежність довжини ртутного стовпчика від температури навколишнього середовища. Функціональні зв'язки притаманні переважно технічним системам і частково – природно-техногенним системам. У суспільних процесах – це зв'язок між елементами розрахункових формул показників – адитивний ( $a + b + c$ ) або мультиплікативний ( $a = bc$ , $c = a/b$ ), а також залежність середніх величин від структури сукупності. Функціональні залежності вивчають точні науки – математика, фізика, хімія. Для дослідження суспільних явищ їх використовують, головним чином, як моделі більш складних стохастичних зв'язків.
<b>Функція щільності розподілу</b>	Характеризує ймовірність потрапляння вибіркового значення випадкової величини в заданий інтервал
<b>Функція розподілу</b>	Характеризує ймовірність значення випадкової величини менше заданого.
<b>Центральні моменти</b>	Є накопиченою сумою відхилень часткових значень випадкової величини від центру розподілу (середнього значення) у певному ступені, віднесеною до стандартного відхилення у тому ж ступені. Використовуються для визначення дисперсії (другий ступінь), асиметрії (третій ступінь), ексцес (четвертий ступінь).
<b>Частота</b>	число повторень окремого значення дискретної випадкової величини, або потрапляння значення неперервної випадкової величини в певний інтервал групування..
<b>Шкали вимірювання параметрів</b>	В географії використовуються чотири шкали кількісної оцінки параметрів географічних об'єктів: 1. <i>Номінальна шкала</i> – використовується для класифікації

	<p>об'єктів за принципом рівності їх властивостей; при цьому число є кодом якісного показника об'єкту. Наприклад, представлення структури народного господарства – 1 – промисловість, 2 – сільське господарство, 3 – сфера послуг і т.д.</p> <p>2. <i>Порядкова шкала</i> – застосовується, коли об'єкти можна розташувати у певному порядку в залежності від зміни будь-якої властивості, але ці зміни не можна оцінити кількісно. Це класифікації, побудовані за схемою високо – середньо – низько, або сильно – середньо – слабо без кількісних критеріїв виділення класів.</p> <p>3. <i>Інтервальна шкала</i> – застосовується, коли інтервали переходу об'єктів із класу в клас рівні і мають кількісно визначені межі, але відсутня точка абсолютного нуля. Прикладом слугують температурні шкали Кельвіна, Цельсія, Фаренгейта, Реомюра тощо.</p> <p>4. <i>Шкала відношень (відносна шкала)</i> – повторює всі властивості інтервальної шкали, але має точку абсолютного нуля (початкову точку відліку координат). Ця шкала є найвищою шкалою вимірювань і застосовується, коли параметри об'єктів мають кількісну оцінку.</p>
--	---

Навчально-методичне видання

Немець Костянтин Аркадійович  
Сегіда Катерина Юріївна

**СТАТИСТИЧНІ МЕТОДИ І ОБРОБКА ГЕОІНФОРМАЦІЇ:**  
навчально-методичний комплекс для самостійної роботи  
студентів, які навчаються за напрямом підготовки «Географія»  
зі спеціальності «Економічна та соціальна географія»

---

Підписано. до друку 04.09.2012. Формат 60×84/16.  
Гарнітура Шкільна. Папір офсетний. Друк різнографічний.  
Ум.-друк. арк. 3,0. Обл.-вид. арк. 2,2. Наклад 50 прим.

Надруковано ФОП Грицак С.Ю., 61103, м. Харків, вул. Дерев'янка 7, к. 86  
Свідоцтво про державну реєстрацію № 04058835Ф0040482 від 01.04.2000  
Тел. (057) 343-75-62, e-mail: ekograf@i.ua